Communication-Free Inter-Operator Interference Management in Shared Spectrum Small Cell Networks

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCCN.2019.2925348, IEEE Transactions on Cognitive Communications and Networking

Communication-Free Inter-Operator Interference Management in Shared Spectrum Small Cell Networks

Cengis Hasan and Mahesh K. Marina

***Abstract*—Emergence of shared spectrum, such as the 3.5 GHz Citizen Broadband Radio Service (CBRS) band in the US, promises to broaden the mobile operator ecosystem and lead to proliferation of small cell deployments. We consider the inter-operator interference problem that arises when multiple small cell networks access the shared spectrum. Towards this end, we take a novel communication-free approach that seeks implicit coordination between operators without explicit communication. The key idea is for each operator to sense the spectrum through its mobiles to be able to model the channel vacancy distribution and extrapolate it for the next epoch. We use reproducing kernel Hilbert space kernel embedding of channel vacancy and predict it by vector-valued regression. This predicted value is then relied on by each operator to perform independent but optimal channel assignment to its base stations taking traffic load into account. Via numerical results, we show that our approach, aided by the above channel vacancy forecasting, adapts the spectrum allocation over time as per the traffic demands and more crucially, yields as good as or better performance than a coordination based approach, even without accounting the overhead of the latter.**

***Index Terms*—Shared spectrum, multi-operator small cell networks, interference prediction and management, machine learning, kernel embedding of distributions**

I. Introduction

Mobile data traffic continues to grow rapidly and scaling the capacity of mobile networks to meet this demand is a key driver for the emerging 5G mobile networks. Making cells smaller and densely deploying them has historically been the biggest contributor to capacity scaling of cellular networks to the extent that they have been named after this concept [1]. However, further increases in cell densification and deployment of small cells are stifled by the current deployment model requiring access to licensed spectrum that is typically possessed by less than a handful of traditional mobile network operators in each country.

Recent regulatory developments in spectrum sharing below 6 GHz[1] that allow sharing of lightly used spectrum held by legacy or public-sector incumbents (e.g., radars and satellite earth stations) via tiered spectrum access models [2], [3] are lowering the barrier for new entrants to the mobile network operator ecosystem by significantly reducing the spectrum acquisition cost. This in turn promises proliferation of small cell deployments, and also allows traditional mobile network operators to offload traffic during peak hours to shared spectrum based small cells. Citizen Broadband Radio Service (CBRS) initiative in the US allowing the shared use of 3.5 GHz band (3500-3700 MHz) via a three-tier access model is a case in point [4]. In the CBRS model, the incumbents make up the top tier. Middle tier users access the spectrum in 10 MHz chunks (up to 70 MHz) using Priority Access Licenses obtained for a medium term (3 years) via auctions. Users in the lowest tier called General Authorized Access (GAA) can access (at least) 80 MHz of spectrum unused by the higher tier users for free. A separate cloud-based management entity called Spectrum Access System (SAS) orchestrated by the regulator ensures that when higher tier users need to use the spectrum they get interference protection from lower tier ones. See [4] for use cases that encourage new entrants to the operator ecosystem leveraging the CBRS style spectrum, which is also expected to rise in amount by an additional 500 MHz with the inclusion of 3.7-4.2 GHz band. Licensed shared access (LSA) model for spectrum sharing that is being promoted for some bands in Europe [5], especially in its dynamic form [6], is another such relevant development. Such shared spectrum use is also now becoming feasible technology-wise with the emergence of MulteFire [7], which brings the high performance and seamless mobile access of LTE to operate solely in unlicensed or shared bands without requiring an anchor in the licensed spectrum.

In this paper, we consider the inter-operator interference management problem[2] that arises when multiple small cell network operators access shared spectrum (e.g., as GAA users in the CBRS 3.5 GHz band). If not carefully managed, interference between multiple operator networks can lead to the "tragedy of the commons" phenomenon, thereby undermining the benefits of shared spectrum, broadened operator ecosystem and dense small cell deployments. However, currently there does not exist a widely accepted approach for operator-level coordination for secondary use of shared spectrum. Even for time-sharing of a given channel, there is no such mechanism like Listen-Before-Talk (LBT) that is mandated as with unlicensed spectrum (e.g., 5 GHz bands used by Wi-Fi networks). While having the cloud-based/centralized SAS mediate access to shared spectrum for interference protection to incumbents and higher tier users is essential, using it to coordinate spectrum sharing among the same tier users (e.g., GAA users in CBRS) as suggested in [8], [9] limits dynamic and fine-grained spectrum use. Another approach would be to have operators exchange their respective spectrum usage information to base their individual interference management

Cengis Hasan is with Faculty of Science, Technology and Communication, University of Luxembourg, email: cengis.hasan@uni.lu

Mahesh K. Marina is with School of Informatics, The University of Edinburgh, UK, email: mahesh@ed.ac.uk

[1]Most existing mobile/wireless communication systems operate below 6 GHz. Even early deployments of 5G mobile networks are expected to use below 6GHz spectrum.

[2]Note that intra-operator interference management is not an issue as the operator can internally coordinate the spectrum allocation among its entities (base stations and mobiles).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCCN.2019.2925348, IEEE Transactions on Cognitive Communications and Networking

and channel selection decisions (e.g., in the spirit of [10]). This latter approach incurs overhead for coordination and may also not be preferable as operators would typically be competitors to each other; moreover, it may be challenging to realize in practice especially if they do not share the underlying infrastructure. We further discuss related work later in Section VII.

Towards this end, *we take a novel communication-free (CF) approach that seeks implicit coordination between operators without explicit communication*. Our approach can be summarized as follows.

- (§III) Operators view the time as a sequence of epochs. In each epoch, every operator through its mobiles senses the spectrum to measure the vacancy of channels in the available spectrum over space. Note that the extent of vacancy of a channel is inversely related to the level of interference on it – low (high) vacancy implies high (low) interference.
- (§IV) The channel vacancy data so obtained from the preceding epochs is used for predicting channel vacancy distribution in the next epoch. Specifically, we use kernel embedding of channel vacancy statistics and learn the evolution of the channel vacancy distribution to the next epoch through an autoregressive (AR) process. This mechanism for predicting channel vacancy distribution constitutes the core contribution of the paper and to our knowledge has not been considered till date in the wireless communications context.
- (§V) The channel vacancy distribution so predicted is used by each operator at the beginning of next epoch for optimally assigning channels to its base stations (BSs) while also taking traffic load into account. This is achieved via the formulation of the optimization problem for communication-free channel assignment for each operator (that leverages the forecasted channel vacancy distribution) and the solution of its Lagrangian relaxation with a sub-gradient descent method.
- (§VI) Via an extensive set of numerical results, we show that our approach forecasts channel vacancy with good accuracy. This in turn is shown to result in spectrum efficiency performance that is as good as or even better than using an ideal inter-operator coordination protocol (i.e., with zero coordination overhead) but also significantly better than channel assignments that do not rely on channel vacancy forecasting.

We now briefly comment on the wider significance of the interference prediction method proposed in this paper. We have considered the particular context of shared spectrum small cell networks as a timely motivating case for our method, it is more generally applicable. Interference prediction in wireless communications would be valuable when making dynamic radio resource control decisions (e.g., spectrum allocation, power control, user scheduling) [49], [50]. As future 5G and beyond mobile networks embrace ultra-dense small cell networks, managing interference would be key to assuring high service quality and this is where interference prediction could enable dynamic resource control and user association decisions [51], [52], [55]. Interference prediction would also be important in other emerging scenarios such as power constrained UAV networks [53]. In terms of spectrum, we have considered the CBRS style shared spectrum as a motivating use case in this

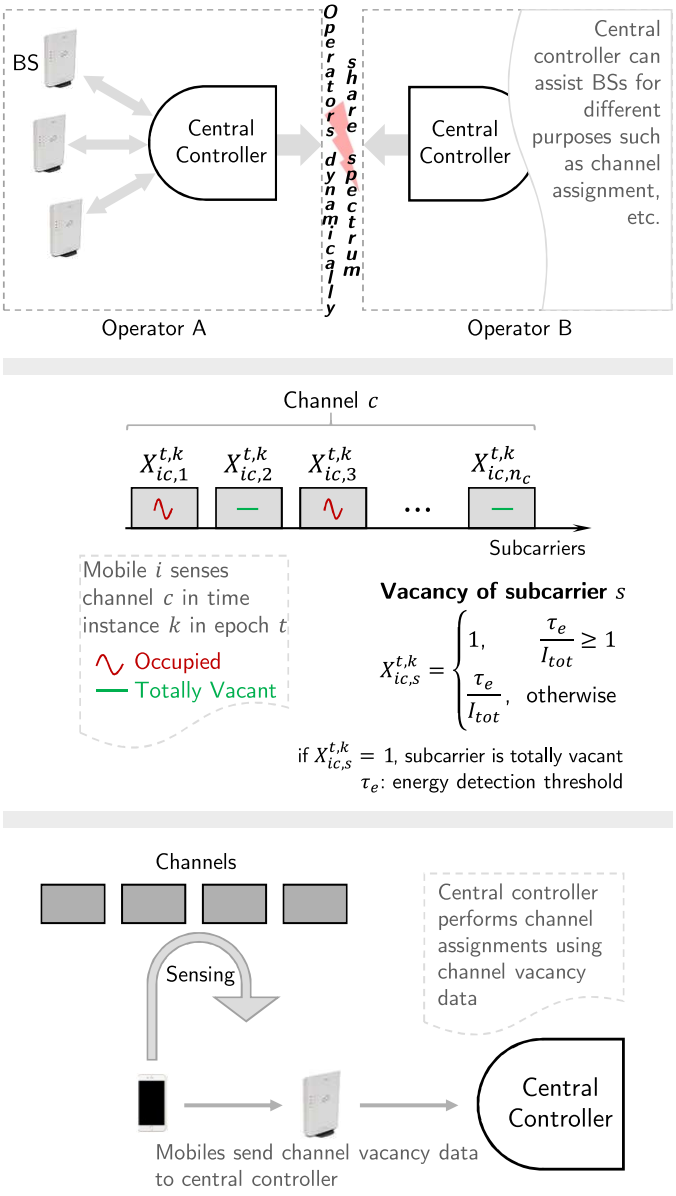


Fig. 1. Illustration of the system model, channel (subcarrier) vacancy measurement at the mobiles and reporting this info to the respective central controllers.

paper keeping in mind that such shared spectrum is already being made available by national regulators for future mobile networks. However, 5G networks are additionally expected to make use of high-frequency millimeter (mmWave) spectrum. Even with such spectrum, the value of spectrum sharing across operators has been highlighted in the literature [54] and thus the framework for communication-free spectrum sharing we present in this paper would also be applicable in the mmWave case.

The next two sections describe the system model and formally state the problem being tackled.

II. System Model

We consider multiple operators, each with its set of small cell BSs deployed in the same environment. For each operator, we assume the presence of a **Central Controller (CC)** for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCCN.2019.2925348, IEEE Transactions on Cognitive Communications and Networking

intra-operator coordination purposes including for channel assignments to BSs as well as to interface with external entities responsible for controlling access to shared spectrum (e.g., SAS in the CBRS context, LSA repository). This is illustrated in Figure 1. Note that there is no explicit coordination between different operators in our model. Moreover, the time in our model is a sequence of epochs.

Shared Spectrum Access. For each operator, as mentioned above, the CC is the entity that interacts with the shared spectrum access management system like SAS and enables access to shared spectrum to the operator. In our model, operators take the role similar to GAA users in CBRS model and can in turn share the available shared spectrum amongst them. The total amount of such shared spectrum can vary over time depending on the activity of the higher tier users and each operator is informed of such changes in availability via its CC. We view the available shared spectrum for multi-operator sharing at any given point in time as a set of channels that we refer to as *Shared Access (SA)* channels.

Spectrum Sensing. We assume that mobile users continually collect interference data by sensing its level in every channel. The sensing can be fine-grained on a subcarrier basis. In that case, the mobile divides the channel into subcarriers and determines the *subcarrier (channel) level vacancy* depending on the level of interference in a subcarrier (and over all subcarriers for a channel) with respect to some energy detection threshold. If it is lower than the threshold, the mobile sets an indicator variable to 1, otherwise it assigns a value which is inversely proportional to the observed interference level. Then, it feeds back this data to its operator's CC through the BSs. By utilizing this data, CC forecasts the interference in the next epoch as detailed in Section IV and accordingly performs channel assignments at the beginning of next epoch, and repeats this process (Figure 1). This results in a stochastic optimization problem as it will become apparent shortly. We assume that sensing is performed periodically throughout an epoch. Such sensing could be implemented in real hardware, for example by simply leveraging the capabilities of mobiles in current (LTE) mobile networks for signal strength measurement and reporting to the associated BS, or following a very efficient algorithm like the one proposed in [13].

Interaction among Operators. Interference on a particular channel is determined by independent channel assignment strategies of different operators and varying channel conditions, and this changes from epoch to epoch. Such a setting makes it challenging for an operator to detect the actions/strategies of other operators, which are "hidden" in the interference data. Therefore, we treat the interference data as *random* and aim to model/predict it for the subsequent epoch.

III. Formal Description

Operators function in a discrete-time setting. Time is viewed as a sequence of time *epochs*, each with *duration* T. There are n_e *number of epochs per day*. We define all the entities of a particular operator without loss of generality: set of BSs as B, set of mobile users as M, set of all *available SA channels* in time epoch t as $C(t)$. A channel c of *bandwidth* w_c is comprised of n_c *subcarriers*.

We assume that each mobile associates with one of the BSs and denote by M_j *the set of users associated with BS* j. For simplicity, we focus on the downlink case; uplink case can be accommodated similarly. We represent by $L_j(t)$ the *total traffic load* to BS j in terms of **bps** in epoch t. The path loss model of transmission between BS j and mobile i is given by $P_{jc,s} g_{jic,s} d_{ji}^{-\alpha}$ where $P_{jc,s}$ and $g_{jic,s}$ are downlink transmission power per subcarrier s and constant path-loss factor (antenna, average channel attenuation) and channel fading coefficient in subcarrier s in channel c, respectively; d_{ji} is distance between j and i, and α is the path loss exponent.

A. Channel Vacancy

Within each epoch t, we assume that mobile i measures a *subcarrier vacancy matrix* given by $\mathbf{X}_{ic}^{t} = [X_{ic,s}^{t,k}] \in [0,1]^{n_c \times n_d}$ where s denotes a subcarrier in the channel c and k denotes a measurement instance in epoch t:

$$X_{ic,s}^{t,k}(\tau_e) = \min\left[1, \frac{\tau_e}{I_{tot}}\right] = \begin{cases} 1, & \frac{\tau_e}{I_{tot}} \geq 1 \\ \frac{\tau_e}{I_{tot}}, & \text{otherwise} \end{cases} \quad (1)$$

meaning that *subcarrier vacancy* $X_{ic,s}^{t,k} \in [0,1]$ shows the ratio of *energy threshold* τ_e and total interference (I_{tot}) in the subcarrier. We set $X_{ic,s}^{t,k}$ to 1 if interference in the subcarrier is below τ_e which can be interpreted to mean that the subcarrier is *vacant*.

Remark 1. *In this work, we choose inversely proportional law to calculate subcarrier vacancy. As an alternative, one can consider any monotonous decreasing function, e.g., exponentially decreasing law* $\exp(-\varrho(\tau_e - I_{tot}))$ *with a tuning parameter* $\varrho(>0)$ *whose higher values favor low discrepancy of* $\tau_e - I_{tot}$.

Consider a particular mobile i and channel c in any one measurement time instance k in epoch t. We define *channel vacancy* in time instance k determined by that mobile as:

$$v_{ic}^{t,k}(\tau_e) = \frac{1}{n_c} \sum_{s=1}^{n_c} X_{ic,s}^{t,k}(\tau_e) \in [0,1], \quad (2)$$

which is basically the average subcarrier vacancy across all subcarriers in the corresponding channel. In the rest of the paper, for the sake of simplicity of exposition, we omit (τ_e) and $v_{ic}^{t,k}$ shall refer to $v_{ic}^{t,k}(\tau_e)$. We assume that variance of vacancy over all subcarriers is low. As such, the study of the case with high variability in vacancy across subcarriers of a channel is out of scope of this paper.

CC uses the channel vacancy data collected from each of its mobiles for modeling/forecasting channel vacancy distribution, which in turn is used for channel assignments to the operator's BSs based on the traffic load of their associated mobiles.

B. Spatial Map of Channel Vacancy

Consider a particular mobile i that is located at ϕ_i in a given measurement time instance and is sensing the vacancy level of each subcarrier s in every channel c, dependent on the effect of surrounding interferers at that time instance (Figure 2). For every measurement instance k in epoch t, CC creates a spatial

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCCN.2019.2925348, IEEE Transactions on Cognitive Communications and Networking

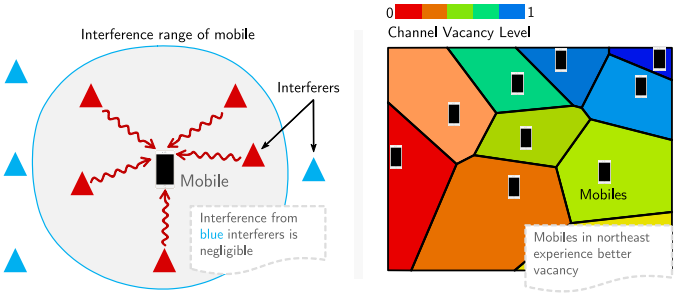


Fig. 2. Left: illustration of interference experienced by a mobile; Right: illustration of spatial channel vacancy level in a particular time instance k of any epoch t – a Voronoi region shows a channel vacancy level.

map of channel vacancy based on data fed back by mobiles. The map guides CC to predict channel vacancy distribution at any location for the upcoming epoch. We model this map at any instance t as a Voronoi tessellation of mobile locations, $\phi_i^{t,k}$ for all $i \in M$, with each region colored based on its channel vacancy level, measured (for the past) or predicted (for the future). To this end, two types of data are fed back by the mobile: experienced channel vacancy and mobile's location, i.e. $\{v_{ic}^{t,k}, \phi_i^{k,t}\}$. This way, in each epoch t, the CC gathers *channel vacancy vector* $[v_{ic}^{t,1}, v_{ic}^{t,2}, \ldots v_{ic}^{t,n_d}]$, where n_d denotes the number of channel measurement instances in each epoch for each mobile i about channel c, using data fed back by the mobiles. Accuracy of the map depends heavily on the number of mobiles and the distribution of their locations. Note that the map changes dynamically based on mobile movements, channel assignments of operators, etc.

Given the above, in our setting, the CC has to predict the channel vacancy distribution for the upcoming epoch t and for any mobile i using: (i) location of mobile i at the beginning of epoch, i.e., $\phi_i^{t,1}$; and (ii) the channel vacancy vector in the Voronoi regions covering location $\phi_i^{t,1}$ in the preceding Δ *days*. More clearly, assume that CC has access to sequences of data for $t' \in \{t-\Delta n_e, t-\Delta n_e+1, \ldots, t-n_e-1, t-n_e\}$ for all $i \in M$ and $c \in C(t)$. For every epoch t by going Δ days backward, CC builds up matrix $\mathbf{V}_{ic}^{<t} = [V_{ic}^{<t,1}, \ldots, V_{ic}^{<t,\Delta}] \in \mathbb{R}^{\Delta \times \Delta}$, for mobile i where $t' = t - (\Delta - z + 1)n_e$:

$$V_{ic}^{<t,z} = \left\{ v_{i'c}^{t',k}, \forall k = 1, \ldots, n_d \middle| \exists i' \in M, \left(\phi_i^{t,1} \in \mathcal{R}_{i'}^{t',k}\right) \right\}, \quad (3)$$

which includes all data in case any mobile i' passed region $\mathcal{R}_{i'}^{t',k}$ in epoch t'. Note that the vacancy map is generated at the granularity of a channel. Also, the time overhead of sensing and feedback of channel vacancy data is relatively lower than the speed of mobiles. Therefore, we suppose that accuracy of channel vacancy map is not affected by mobility.

C. Sensing and Feedback Overhead

Time cost to sense and process vacancy data for a subcarrier is denoted by $\bar{T}_S$. Total time cost of doing it for all channels can be given by $\bar{T}_S \sum_{c \in C(t)} n_c$. Besides, there is a time cost of feeding back channel vacancy and location data to the BS which are denoted by $\bar{T}_F^V$ and $\bar{T}_F^L$. Thus, total channel vacancy feedback time cost is given by $\bar{T}_F^V |C(t)|$. Note that both $\bar{T}_F^V$ and $\bar{T}_L^V$ depend heavily on block size (number of bits) of channel vacancy, location data and feedback transmission frequency. Location data is not needed to be fed back always unless mobile moves. If mobile does not move during the epoch, then the corresponding location related feedback cost will be zero. Otherwise, mobile will send its location data to its associated BS with cost $\bar{T}_L^V$. This overall cost limits n_d and given by:

$$n_d \leq \left\lfloor \frac{T}{\bar{T}_S \sum_{c \in C(t)} n_c + \bar{T}_F^V |C(t)| + \bar{T}_F^L} \right\rfloor. \quad (4)$$

n_d needs to be chosen as high as possible in order to capture the statistical properties of channel vacancy.

There exists mechanisms in the literature (e.g., [13]) that enable efficient and fast sensing. For accurate channel sensing in the presence of noise, multiple readings are needed. With a 20MHz ADC and 1000 energy readings, sensing time will be 0.05ms. With a 20MHz channel and 1200 subcarriers, the time needed to sense a subcarrier will be about $0.05/1200$ ms ≈ 4.17 μs. Moreover, by setting the energy detection threshold much higher than the noise power level, the effects of uncertainties around estimation of noise power level can be kept minimal and quick reliable channel sensing enabled.

Errors in sensing and feedback can occur but we do not explicitly address them in the paper. As such, they fall outside the scope of this paper. However, we can account for the discrepancy between extrapolated $\hat{\bar{v}}_{ic}$ and observed $\bar{v}_{ic}$ channel vacancy data. Specifically, we can define percentage error of prediction as $100|\bar{v}_{ic} - \hat{\bar{v}}_{ic}|/\bar{v}_{ic}$. If it is below a certain threshold, then we may not trust the extrapolated data for coming epoch; instead the observed data from previous epoch can be used for performing channel assignments.

TABLE I
LIST OF NOTATIONS

Parameter	Definition
$C(t)$	available channels in epoch t
w_c, n_c	bandwidth and number of subcarriers of channel c
$L_j(t)$	total traffic load to BS j in epoch t in **bps**
τ_e	energy threshold
$v_{ic}^{t,k}$	vacancy of channel c at mobile i at time instance k in epoch t
$\phi_i^{t,k}$	location of mobile i at time instance k in epoch t
n_d	number of channel measurement instances in each epoch
n_e	number of epochs per day
Δ	number of preceding days whose data is used for prediction
$\mu_{\mathcal{D}}$	kernel mean embedding of channel vacancy distribution $\mathcal{D}$
$\kappa(\cdot,\cdot)$	kernel function
ρ	order of AR model
λ	smoothing parameter in least-squares functional
$\hat{R}_{jic}$	extrapolated raw throughput from BS j to mobile i in channel c

D. Statistics of Channel Vacancy Data

There are several phenomena that change the statistics of channel vacancy data:

- **Channel assignment**: operators may change channel assignments to optimize the service to their respective mobiles.
- **Dynamic channel conditions**: wireless transmission channel may vary within an epoch.
- **Mobility**: user nodes may move.
- **Power control**: downlink transmission power may be adapted again to better serve users.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCCN.2019.2925348, IEEE Transactions on Cognitive Communications and Networking

Probability distribution of channel vacancy contains all these dimensions. However it would be impossible to distinguish all these aspects from channel vacancy data. Therefore, we use a holistic framework that views this data as a source to predict its distribution for the upcoming epoch.

Definition 1 (Subcarrier Vacancy Probability)**.** *For any threshold τ_e and $\tilde{N}$ interferers, it is defined as*

$$\mathbb{P}[X_{ic,s} \leq x] = 1 - \mathbb{P}\left[x \leq \frac{\tau_e}{\sum_{j\in\tilde{N}} I_{jic,s}}, x < 1\right] \tag{5}$$

Lemma 1. *Subcarrier vacancy probability is given by*

$$\mathbb{P}[X_{ic,s} \leq x] = 1 - \tilde{p}_{ic,s}\left(\frac{\tau_e}{x}\right)\mathbf{1}\{x < 1\}. \tag{6}$$

Proof. We can write

$$\mathbb{P}\left[x \leq \frac{\tau_e}{\sum_{j\in\tilde{N}} I_{jic,s}}, x < 1\right] = \mathbb{P}\left[\sum_{j\in\tilde{N}} I_{jic,s} \leq \frac{\tau_e}{x}\right]\mathbf{1}\{x < 1\},$$

where $\mathbf{1}\{\cdot\}$ is indicator function, and

$$\mathbb{P}\left[\sum_{j\in\tilde{N}} I_{jic,s} \leq \frac{\tau_e}{x}\right] = \tilde{p}_{ic,s}\left(\frac{\tau_e}{x}\right). \tag{7}$$

Thus, we obtain the result given by (6). □

Similarly, we can calculate the expected value and variance of subcarrier vacancy using cumulative distribution function as following:

$$\bar{X}_{ic,s} = \int_0^\infty \mathbb{P}[X_{ic,s} \geq x]dx = \int_0^1 \tilde{p}_{ic,s}\left(\frac{\tau_e}{x}\right)dx, \tag{8}$$

$$\begin{aligned}\sigma^2_{X_{ic,s}} &= 2\int_0^\infty x\mathbb{P}[X_{ic,s} \geq x]dx - \left(\int_0^\infty \mathbb{P}[X_{ic,s} \geq x]dx\right)^2 \\ &= 2\int_0^1 x\tilde{p}_{ic,s}\left(\frac{\tau_e}{x}\right)dx - \left(\int_0^1 \tilde{p}_{ic,s}\left(\frac{\tau_e}{x}\right)dx\right)^2. \end{aligned} \tag{9}$$

Rayleigh Fading: We calculate expected value and variance of subcarrier vacancy considering Rayleigh fading model. Note that interference $I_{jic,s}$ follows exponential distribution; thus, we have

$$\tilde{p}_{ic,s}\left(\frac{\tau_e}{x}\right) = \left[\prod_{j\in\tilde{N}} \frac{1}{\bar{I}_{jic,s}}\right]\sum_{j\in\tilde{N}} \frac{e^{-\frac{\tau_e}{x\bar{I}_{jic,s}}}}{\prod_{k\in\tilde{N}\setminus j}\left(\frac{1}{\bar{I}_{kic,s}} - \frac{1}{\bar{I}_{jic,s}}\right)}$$

which results in the following:

Expected Value and Variance of Subcarrier Vacancy: Rayleigh Fading

$$\bar{X}_{ic,s} = 1 - \sum_{j\in\tilde{N}} \mathcal{I}(j)\left(e^{-\frac{\tau_e}{\bar{I}_{jic,s}}} + \frac{\tau_e}{\bar{I}_{jic,s}}\mathrm{Ei}\left(-\frac{\tau_e}{\bar{I}_{jic,s}}\right)\right)$$

and

$$\sigma^2_{X_{ic,s}} = 1 - \sum_{j\in\tilde{N}} \mathcal{I}(j)\left(e^{-\frac{\tau_e}{\bar{I}_{jic,s}}}\left(1 - \frac{\tau_e}{\bar{I}_{jic,s}}\right) - \frac{\tau_e^2}{\bar{I}^2_{jic,s}}\mathrm{Ei}\left(-\frac{\tau_e}{\bar{I}_{jic,s}}\right)\right) - \bar{X}^2_{ic,s}$$

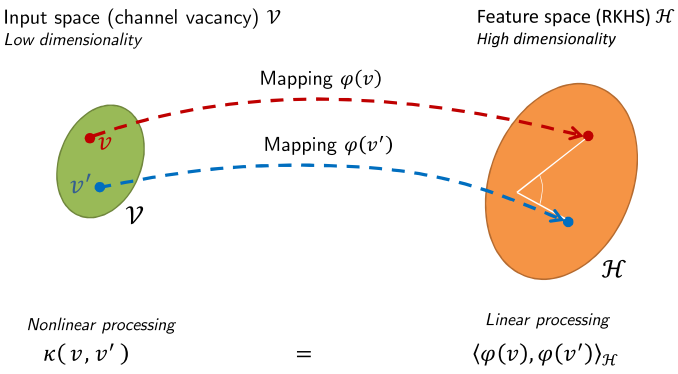


Fig. 3. Reproducing kernel Hilbert space illustration.

where

$$\mathcal{I}(j) = \frac{\prod_{k\in\tilde{N}\setminus j}\frac{1}{\bar{I}_{kic,s}}}{\prod_{k\in\tilde{N}\setminus j}\left(\frac{1}{\bar{I}_{kic,s}} - \frac{1}{\bar{I}_{jic,s}}\right)}$$

Definition 2 (Channel Vacancy Probability)**.** *For any threshold τ_e, it can be given by*

$$\mathbb{P}[v_{ic} \leq \chi] = \mathbb{P}\left[\frac{1}{n_c}\sum_{s=1}^{n_c} X_{ic,s} \leq \chi\right]. \tag{10}$$

Theorem 1. *(From central limit theorem): For large values of n_c, i.e, when $n_c \to \infty$, probability density of channel vacancy is given by normal distribution:*

$$\mathcal{D}_{ic} \to \mathsf{pdf}_{v_{ic}}(\chi) = \frac{1}{\sqrt{2\pi\sigma^2_{v_{ic}}}}\exp\left(-\frac{(\chi - \bar{v}_{ic})^2}{\sigma^2_{v_{ic}}}\right) \tag{11}$$

where expected value and variance of v_{ic} denoted by $\bar{v}_{ic}$ and $\sigma^2_{v_{ic}}$ are given by

$$\bar{v}_{ic} = \frac{1}{n_c}\sum_{s=1}^{n_c}\bar{X}_{ic,s} \text{ and } \sigma^2_{v_{ic}} = \frac{1}{n_c^2}\sum_{s=1}^{n_c}\sigma^2_{X_{ic,s}}, \tag{12}$$

respectively.

Proof. See appendix. □

IV. Reproducing Kernel Hilbert Space Embedding of Channel Vacancy Data and AR Models

We use *vector-valued regression* for learning how the (embedded) distribution of channel vacancy evolves from one epoch to the next epoch. The training set data consists of vectors of channel vacancy data corresponding to a number of epochs. We want to predict channel vacancy data for the next epoch, which also is in the vector form. Therefore, the type of learning of interest here is vector-valued regression.

Channel vacancy levels are sampled i.i.d. from the respective *distributions*, $\mathcal{D}_{ic}^{t-\Delta n_e}, \mathcal{D}_{ic}^{t-(\Delta-1)n_e}, \ldots, \mathcal{D}_{ic}^{t-n_e}$. We aim to construct a distribution $\hat{\mathcal{D}}_{ic}^t$ to be as close as possible to the as yet unobserved $\mathcal{D}_{ic}^t$. Then, we calculate extrapolated expected value of channel vacancy $\bar{v}_{ic}^t$ for the upcoming epoch t.

In any epoch t, distribution $\mathcal{D}_{ic}^t$ is determined by vacancy probabilities of subcarriers $\tilde{p}_{ic,s}$. Note that $\tilde{p}_{ic,s}$ is affected by several factors: number of interferers (which is in turn

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCCN.2019.2925348, IEEE Transactions on Cognitive Communications and Networking

dependent on the channel assignment strategies of operators), power per subcarrier, distance, channel fading coefficient, and other physical phenomena. The expected value and variance of a distribution randomly fluctuates in every epoch. So our aim is to infer any underlying *probabilistic* relation or similarity between epochs. From the previous section, we have that channel vacancy follows Gaussian distribution (when n_c is high). However, the mean of channel vacancy would be very chaotic due to the non-stationary environment with multiple operators adapting their channel assignments reacting to each others' actions. So, we need a technique which does not require any prior knowledge about distributions.

A. Reproducing Kernel Hilbert Spaces

In the previous two sections, we have described the system model and the problem setting. We now present the mathematical technique we employed for prediction of channel vacancy distribution. Formally, a Hilbert space is an abstract vector space possessing the structure of an inner product. It allows length and angle to be measured. So, we have linear processing in Hilbert spaces with low computational complexity. The input space, denoted by $\mathcal{V}$, in our case is the channel vacancy data. It is only a scalar between 0 and 1. Therefore, we say that it has low dimensionality. We call Hilbert space as *feature space* and has higher dimensionality (at least the space of vectors). To do linear operations, we have to map every vacancy data to Hilbert space. It is done by *mapping* φ, which has scalar input and vector output. In general, it is challenging to find such mappings. Therefore, we restrict ourselves to those Hilbert spaces which are called as *reproducing kernel Hilbert spaces* (RKHS) (See Fig. 3).

Any RKHS $\mathscr{H}$ with *kernel* κ is a Hilbert space of functions $\mathcal{V} \to \mathbb{R}$ with dot product $\langle \cdot, \cdot \rangle_{\mathscr{H}}$, satisfying the *reproducing property*:

$$\langle \varphi(v), \varphi(v') \rangle_{\mathscr{H}} = \langle \kappa(v, \cdot), \kappa(v', \cdot) \rangle_{\mathscr{H}} \tag{13}$$

$$\langle \kappa(v, \cdot), \kappa(v', \cdot) \rangle_{\mathscr{H}} = \kappa(v, v'), \tag{14}$$

$$\langle f(\cdot), \kappa(v, \cdot) \rangle_{\mathscr{H}} = f(v) \tag{15}$$

where $v, v' \in \mathcal{V}$. The notation $f(\cdot)$ refers to the function itself in the abstract. The linear map from a function f on $\mathcal{V}$ to its value at v can be seen as an inner product. $\kappa(v, v')$ can be interpreted as a *non-linear similarity measure* between v and v'.

In the following, we introduce and solve three different autoregressive (AR) models related to channel vacancy data and its expected value.

B. Kernel Mean Embedding of Channel Vacancy

For any $\mathcal{D}$, kernel mean embedding of channel vacancy v is given by the following mappings:

$$\mu_{\mathcal{D}} := \mathbb{E}_{v' \sim \mathcal{D}}[\kappa(v', \cdot)] \quad \text{and}$$

$$\hat{\mu}_{\mathcal{D}} := \frac{1}{n_d} \sum_{k=1}^{n_d} \kappa(v^k, \cdot) \quad \text{(empirical estimation)}$$

Whenever sufficient condition $\mathbb{E}_v[\kappa(v, v')] < \infty$ is met, then reproducing property imposes $\langle \hat{\mu}_{\mathcal{D}}, f \rangle_{\mathscr{H}} = \frac{1}{n_d} \sum_{k=1}^{n_d} f(v^k)$.

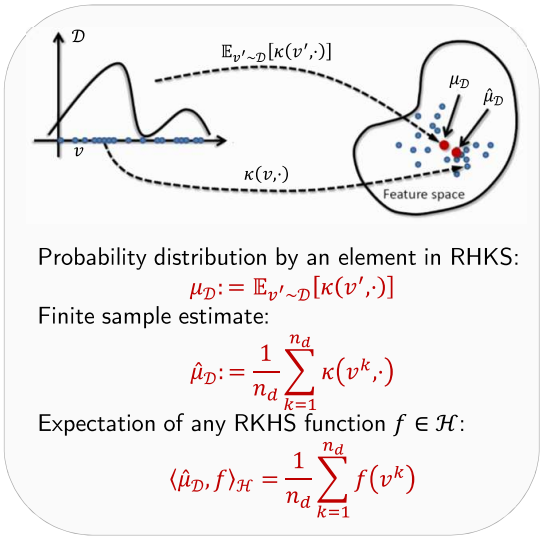


Fig. 4. Illustration of kernel mean embedding of channel vacancy.

We refer to [14], [15] for further reading. In Fig. 4, there is an illustration of the above concepts. Without loss of generality, we consider a particular mobile and channel at the beginning of an epoch t with channel vacancy vector[3] V^z for $z = 1, \ldots, \Delta$.

C. AR Model of Kernel Mean Embedding of Channel Vacancy (KME)

We suppose that channel vacancy distributions between time epochs can be approximated by an AR process of kernel embedding means, i.e.,

$$\hat{\mu}_{z+1} = \sum_{r=1}^{\rho} \mathbf{\Lambda}_r \hat{\mu}_{z-r+1} + \epsilon_z \,^4$$

for some operator $\mathbf{\Lambda}_r : \mathscr{H} \to \mathscr{H}$, for all $r = 1, \ldots, \rho$ such that ϵ_z for all $z = 1, \ldots, \Delta - 1$ are independent zero-mean random variables. Any operator $\mathbf{\Lambda}_r \in \mathscr{L}$ is a *linear* operator where $\mathscr{L}$ represents a space of *linear operators* as defined in [17]. Finding which operator is suitable is the main question here, and we fortunately are able to learn it by solving following least-squares functional with *smoothing parameter* $\lambda > 0$:

$$\min_{\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_\rho} \sum_{z=\rho}^{\Delta-1} \left\| \hat{\mu}_{z+1} - \sum_{r=1}^{\rho} \mathbf{\Lambda}_r \hat{\mu}_{z-r+1} \right\|_{\mathscr{H}}^2 + \lambda \sum_{r=1}^{\rho} \|\mathbf{\Lambda}_r\|_{\mathscr{L}}^2. \tag{16}$$

Theorem 2. *For* $r = 1, \ldots, \rho$*, solution of* (16) *is given by*

$$\hat{\mathbf{\Lambda}}_r = \sum_{r'=1}^{\rho} \Upsilon_{r,r'} \tilde{\mathbf{m}}_{r'} = \sum_{z,z'=\rho}^{\Delta-1} Q_{z,z'} \hat{\mu}_{z+1} \sum_{r'=1}^{\rho} \Upsilon_{r,r'} \hat{\mu}_{z'-r'+1}^{\top} \tag{17}$$

Proof. See appendix. □

If we apply learned operators $\hat{\mathbf{\Lambda}}$ to the last observed data from $\Delta - \rho + 1$ to Δ, then the result is a prediction of kernel mean embedding in $\Delta + 1$ with $\hat{\mu}_{\Delta+1} = \sum_{r=1}^{\rho} \hat{\mathbf{\Lambda}}_r \hat{\mu}_{\Delta-r+1}$ which

[3]Note that V^z is same as $V^{<t,z}$ – we omit $< t$ for brevity.
[4]We omit $\mathcal{D}$ to simplify presentation, i.e. μ_z refers to $\hat{\mu}_{\mathcal{D}^z}$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCCN.2019.2925348, IEEE Transactions on Cognitive Communications and Networking

we expect to approximate unknown $\mu_{\Delta+1}$. Note that $\hat{\mu}_{\Delta+1}$ can be further calculated by a weighted linear combination of observed distributions:

$$\hat{\mu}_{\Delta+1} = \sum_{z,z'=\rho}^{\Delta-1} Q_{z,z'} \hat{\mu}_{z+1} \sum_{r,r'=1}^{\rho} \Upsilon_{r,r'} \underbrace{\hat{\mu}_{z'-r'+1}^{\top} \hat{\mu}_{\Delta-r+1}}_{\langle \hat{\mu}_{z'-r'+1}, \hat{\mu}_{\Delta-r+1} \rangle_{\mathscr{H}}}. \quad (18)$$

We can now compute expected value of any function $f \in \mathscr{H}$ by using predicted $\hat{\mu}_{\Delta+1}$ with weighted linear combinations of f at channel vacancy data: $\hat{\mathbb{E}}[f(v)] = \langle \hat{\mu}_{\Delta+1}, f \rangle_{\mathscr{H}}$

$$\begin{aligned}\hat{\mathbb{E}}[f(v)] &= \sum_{z,z'=\rho}^{\Delta-1} Q_{z,z'} \langle \hat{\mu}_{z+1}, f \rangle_{\mathscr{H}} \sum_{r,r'=1}^{\rho} \Upsilon_{r,r'} \langle \hat{\mu}_{z'-r'+1}, \hat{\mu}_{\Delta-r+1} \rangle_{\mathscr{H}} \\ &= \sum_{z=\rho}^{\Delta-1} \sum_{k=1}^{n_d} \frac{b_z}{n_d} f(v^{k,z+1}) \quad (19)\end{aligned}$$

with $b_z = \sum_{z'=\rho}^{\Delta-1} Q_{z,z'} \sum_{r,r'=1}^{\rho} \Upsilon_{r,r'} \langle \hat{\mu}_{z'-r'+1}, \hat{\mu}_{\Delta-r+1} \rangle_{\mathscr{H}}$.

D. Kernel Embedding of AR Model of Channel Vacancy (KEC)

In this model, we consider that kernel embedding of channel vacancy which itself follows an AR model. We implement results from [16] Yule-Walker equations as in RKHS where channel vacancy values are mapped from the input space to RKHS using nonlinear transformation. Define nonlinear map $\varphi(\cdot)$ from input space to RKHS $\mathscr{H}$. So, each channel vacancy data v^z is mapped to its corresponding $\varphi(v^z)$. AR process in RKHS $\mathscr{H}$ is given by

$$\varphi(v^{z+1}) = \sum_{r=1}^{\rho} \ell_r \varphi(v^{z-r+1}) + \varphi(\epsilon_z) \quad (20)$$

assuming that $\varphi(\epsilon_z)$ is uncorrelated with $\varphi(v^{z-r+1})$, for all r. By applying Yule-Walker equations, one can find *model parameters* $\hat{\ell}_1, \ldots, \hat{\ell}_\rho$, which are calculated by solving following problem: for $r = 1, \ldots, \rho$,

$$\mathbb{E}[\tilde{\kappa}(v^z, v^{z-r+1})] = \sum_{r'=1}^{\rho} \ell_{r'} \mathbb{E}[\tilde{\kappa}(v^{z-r'+1}, v^{z-r+1})] \quad (21)$$

where $\tilde{\kappa}(\cdot,\cdot)$ is *centered version* of kernel $\kappa(\cdot,\cdot)$ and its expectation is defined with

$$\begin{aligned}\mathbb{E}[\tilde{\kappa}(v^z, v^{z'})] = \frac{1}{n_d^2 \Delta} \Bigg[&\sum_{k,k'=1}^{n_d} \kappa(v^{k,z}, v^{k',z'}) - \sum_{z''=1}^{\Delta} \sum_{k,k'=1}^{n_d} \kappa(v^{k,z}, v^{k',z''}) \\ &- \sum_{z''=1}^{\Delta} \sum_{k,k'=1}^{n_d} \kappa(v^{k,z'}, v^{k',z''}) + \frac{1}{\Delta} \sum_{z'',z'''=1}^{\Delta} \sum_{k,k'=1}^{n_d} \kappa(v^{k,z''}, v^{k',z'''}) \Bigg]. \quad (22)\end{aligned}$$

In [16], solution of (21) is shown to be

$$\hat{\ell}_r = \sum_{r'=1}^{\rho} [W^{-1}]_{r,r'} \mathbb{E}[\tilde{\kappa}(v^{\Delta}, v^{\Delta-r})], \quad (23)$$

where $[W^{-1}]_{r,r'} \in \mathbf{W}^{-1}$, and $\mathbf{W} \in \mathbb{R}^{\rho \times \rho}$ with entries $W_{r,r'} = \mathbb{E}[\tilde{\kappa}(v^{\Delta}, v^{\Delta+r-r'})]$.

Pre-image Problem: Once model parameters $\hat{\ell}_1, \ldots, \hat{\ell}_\rho$ are found, we can predict

$$\hat{\varphi}(v^{z+1}) = \sum_{r=1}^{\rho} \hat{\ell}_r \varphi(v^{z-r+1}),$$

where $\hat{\varphi}(v^{z+1})$ lies in the feature space (i.e. RKHS $\mathscr{H}$). However, we need to map back $\hat{\varphi}(v^{z+1})$ from feature space to the input space, where one identifies the best $\hat{v}^{z+1}$, called as pre-image, in the input space whose image $\varphi(v^{z+1})$ is as close as possible to $\hat{\varphi}(v^{z+1})$.

In this work, we consider only radial kernels which enables to find pre-image by finding the optimum of (see [16])

$$\hat{v}^{z+1} = \arg\max_{v} \sum_{r=1}^{\rho} \hat{\ell}_r \kappa(v, v^{\Delta-r+1}).$$

By taking first derivative with respect to v one can extrapolate the expected value of channel vacancy for epoch t which corresponds to solve following fixed-point equation:

$$\sum_{r=1}^{\rho} \hat{\ell}_r \mathbb{E}_v[(v - v^{\Delta-r+1}) \kappa(v, v^{\Delta-r+1})] = 0. \quad (24)$$

Lemma 2. *Gaussian kernel with variance σ_κ^2 results in fixed-point equation*

$$\sum_{r=1}^{\rho} \hat{\ell}_r \mathcal{J}_r(v^{\Delta-r+1}) = 0$$

where $\mathcal{J}_r(v^{\Delta-r+1}) = \mathbb{E}_v[(v - v^{\Delta-r+1}) \kappa(v, v^{\Delta-r+1})]$ is given by

$$\begin{aligned}&\frac{\sigma_\kappa^3 (v^{\Delta-r+1} - \bar{v})}{2(\sigma_v^2 + \sigma_\kappa^2)^{3/2}} e^{-\frac{(v^{\Delta-r+1} - \bar{v})^2}{2(\sigma_v^2 + \sigma_\kappa^2)}} \left(\operatorname{erf}\left(\frac{(v^{\Delta-r+1} - 1)\sigma_v}{\sqrt{2\sigma_\kappa^2(\sigma_v^2 + \sigma_\kappa^2)}} + \frac{(\bar{v} - 1)\sigma_\kappa}{\sqrt{2\sigma_v^2(\sigma_v^2 + \sigma_\kappa^2)}} \right) \right. \\ &\left. - \operatorname{erf}\left(\frac{\sigma_v v^{\Delta-r+1}}{\sqrt{2\sigma_r^2(\sigma_v^2 + \sigma_\kappa^2)}} + \frac{\bar{v}\sigma_\kappa}{\sqrt{2\sigma_v^2(\sigma_v^2 + \sigma_\kappa^2)}} \right) \right) \\ &+ \frac{\sigma_v \sigma_\kappa^2}{\sqrt{2\pi}(\sigma_v^2 + \sigma_\kappa^2)} \left(e^{-\frac{(\bar{v}-1)^2}{2\sigma_{\bar{v}}^2} - \frac{(v^{\Delta-r+1}-1)^2}{2\sigma_\kappa^2}} - e^{-\frac{\bar{v}^2}{2\sigma_{\bar{v}}^2} - \frac{(v^{\Delta-r+1})^2}{2\sigma_\kappa^2}} \right) \quad (25)\end{aligned}$$

Proof. From Theorem 1, we have that channel vacancy has normal distribution with mean $\bar{v}$ and variance σ_v. Then, we have

$$\mathcal{J}_r(v^{\Delta-r+1}) = \int_0^1 \frac{v - v^{\Delta-r+1}}{\sqrt{2\pi\sigma_v^2}} \exp\left(-\frac{(v - v^{\Delta-r+1})^2}{2\sigma_\kappa^2} \right) \exp\left(-\frac{(v-\bar{v})^2}{2\sigma_v^2} \right) dv$$

of which solution can be shown to be eq. (25). □

Empirical Estimation: Fixed-point equation can have the following form:

$$\hat{v}^{k,t} = \frac{\sum_{r=1}^{\rho} \hat{\ell}_r \kappa(\hat{v}^{k,t}, v^{k,\Delta-r+1}) v^{k,\Delta-r+1}}{\sum_{r=1}^{\rho} \hat{\ell}_r \kappa(\hat{v}^{k,t}, v^{k,\Delta-r+1})}, \quad (k = 1, \ldots, n_d). \quad (26)$$

Then, we calculate empirical mean of channel vacancy forecast by $\hat{\bar{v}}^t = \frac{1}{n_d} \sum_{k=1}^{n_d} \hat{v}^{k,t}$.

In case of Gaussian kernel with variance σ_κ^2, note that optima of $\sum_{r=1}^{\rho} \hat{\ell}_r \exp(-(v^{k,t} - v^{k,\Delta-r+1})^2 / 2\sigma_\kappa^2)$ are mostly affected by $v^{k,\Delta-r+1}$, $r = 1, \ldots, \rho$. Therefore, we initialize fixed-point algorithm for each $r = 1, \ldots, \rho$ from $v^{k,\Delta-r+1}$, and select the maximum accordingly.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCCN.2019.2925348, IEEE Transactions on Cognitive Communications and Networking

E. AR Model of Kernel Embedding of Channel Vacancy Expected Value (KEV)

We model the expected value of channel vacancy as a AR model in RKHS $\mathscr{H}$. Note that we only embed expected value of channel vacancy in RKHS. Then, we have $\kappa(\bar{v},\cdot)$ which maps expected value of channel vacancy $\bar{v}$ to RKHS $\mathscr{H}$. AR process of such a mapping is given by $\kappa(\bar{v}^{z+1},\cdot) = \sum_{r=1}^{\rho} \ell_r \kappa(\bar{v}^{z-r+1},\cdot) + \epsilon_z$. Least-squares formulation to calculate parameters $\ell_1,\ldots,\ell_\rho$:

$$\min_{\ell_1,\ldots,\ell_\rho} \sum_{z=\rho}^{\Delta-1} \left\| \kappa(\bar{v}^{z+1},\cdot) - \sum_{r=1}^{\rho} \ell_r \kappa(\bar{v}^{z-r+1},\cdot) \right\|_{\mathscr{H}}^2 + \lambda \sum_{r=1}^{\rho} \|\ell_r\|_{\mathscr{L}}^2, \tag{27}$$

where $\ell_r \in \mathscr{L}$, for all r, are vector regressors that can be calculated by $\hat{\ell}_r = \sum_{z,z'=\rho}^{\Delta-1} \tilde{Q}_{z,z'} \bar{v}^{z+1} \sum_{r'=1}^{\rho} [\tilde{M}^{-1}]_{r,r'} \bar{v}^{z'-r'+1}$, for all $r = 1,\ldots,\rho$, where $\tilde{Q}_{z,z'} \in \tilde{\mathbf{Q}} = (\tilde{\mathbf{K}} + \lambda \mathbf{I})^{-1}$, and entries $\tilde{\mathbf{K}} \in \mathbb{R}^{(\Delta-1)\times(\Delta-1)}$ are given by $K_{z,z'} = \kappa(\bar{v}^z, \bar{v}^{z'})$; $[\tilde{M}^{-1}]_{r,r'} \in \tilde{\mathbf{M}}^{-1}$ where $\tilde{\mathbf{M}} \in \mathbb{R}^{\rho\times\rho}$ with entries

$$\tilde{M}_{r,r'} = \sum_{z,z'=\rho}^{\Delta-1} \tilde{Q}_{z,z'} \kappa(\bar{v}^{z-r'+1}, \bar{v}^{z'-r+1}).$$

Note that this is a direct result from Theorem (2) where instead of calculating kernel mean embedding $\hat{\mu}$, we only need to calculate kernel value $\kappa(\cdot,\cdot)$ of expected values of channel vacancies.

Remark 2. *Note that KME and KEV are methods in which we first do kernel embedding of some data and then use AR model in RKHS. On the other hand, KEC defines AR model of channel vacancy data in input space, then embeds it to RKHS which results in another AR model. So in KEC, error term in input space is also embedded to RKHS. Whereas in KME and KEV, error is only defined in RKHS.*

Remark 3. *Although we focus on interference prediction in the context of multi-operator shared spectrum small-cell networks, kernel embedding technique employed is general and can be applied to any wireless communications context where interference prediction is of interest.*

V. Channel Assignment

In this section, we study channel assignment for shared spectrum small cell networks. Without loss of generality, we define problem parameters of a particular operator. We examine two cases:

- *Proposed Communication-Free (CF) Scheme* in which channel assignments are simultaneously made by each operator at the beginning of each epoch. Extrapolation of channel vacancy to the next epoch is essential in this case.
- *Inter-Operator Coordination Protocol (CP) based approach*: With this approach, operators asynchronously perform assignments at the beginning of the epoch. Once an operator performs channel assignment for the current epoch, it informs other operators. Essentially each operator performs its assignments taking into account the known assignments from other operators. This reflects the approach adopted in [10]. Note that we consider an idealized version assuming that zero coordination overhead is incurred.

A. Proposed Communication-Free (CF) Scheme

We assume that operators determine their channel assignment strategies without communicating with the others and channel vacancy distribution is extrapolated as introduced in the previous section. However, we assume that channel assignments are made instantaneously at the beginning of each epoch by all operators *tuned to a global time*.

Maximum bandwidth that can be allocated to a mobile $i \in M_j$ can be given by $\sum_{c\in C(t)} w_c/|M_j|$; moreover, we are able to calculate average interference using extrapolated channel vacancy data, for example, any mobile i in channel c shall experience $\tau_e/\bar{v}_{ic}$ interference, where note that if $\bar{v}_{ic} = 1$ then, interference is less than or equal to τ_e which shall be considered as negligible. Assignment variable is given by

$$\mathsf{x}_{cjj'} = \begin{cases} 1, & \text{channel } c \text{ is assigned to BS } j \text{ and } j', \\ 0, & \text{otherwise.} \end{cases} \tag{28}$$

where $\mathsf{x}_{cjj} \equiv \mathsf{x}_{cj}$ basically means that channel c is assigned to BS j. Besides, if channel c is assigned to BSs j and j', then we have $\mathsf{x}_{cjj'} = \mathsf{x}_{cj}\mathsf{x}_{cj'}$ which can be linearized with

$$\mathsf{x}_{cjj'} \le \mathsf{x}_{cj}, \quad \forall c \in C(t), \forall j, j' \in B,$$
$$\mathsf{x}_{cj} + \mathsf{x}_{cj'} - \mathsf{x}_{cjj'} \le 1, \quad \forall c \in C(t), \forall j, j' \in B.$$

We use Jensen's inequality to define a lower bound for throughput which requires to know only expected value of channel vacancy for calculating channel assignments. We define *extrapolated average raw throughput*[5] of mobile i in channel c as following:

$$\begin{aligned} \hat{R}_{jic}(\mathbf{x}) &= \frac{w_c}{|M_j|} \log\left(1 + \frac{P_{jc}\bar{g}_{jic} d_{ji}^{-\alpha}}{\tau_e/\hat{\bar{v}}_{ic} + N_0}\right) \mathsf{x}_{jc} \\ &\ge \hat{\mathbb{E}}_{v_{ic}} \left[\frac{w_c}{|M_j|} \log\left(1 + \frac{P_{jc}\bar{g}_{jic} d_{ji}^{-\alpha}}{\tau_e/v_{ic} + N_0}\right) \mathsf{x}_{jc} \right] \end{aligned} \tag{29}$$

where $P_{jc} = \frac{1}{n_c}\sum_{s=1}^{n_c} P_{jc,s}$, $\bar{I}_{jic} = \frac{1}{n_c}\sum_{s=1}^{n_c} I_{jic,s}$ is average interference across subcarriers resulting from BS j to mobile i. Note that intra-operator interference is hidden in estimated $\tau_e/\hat{\bar{v}}_{ic}$. Moreover, we may set an upper bound I_{max} (e.g. $I_{max} = \tau_e$) for intra-operator interference that a mobile receives, i.e., $\sum_{j'\in B\setminus j} \bar{I}_{j'ic}\mathsf{x}_{cjj'} \le I_{max}, \forall c \in C(t)$. Lower bound of raw throughput is chosen to be the constraint related to the traffic accumulated in a BS. For every BS j and traffic load $L_j(t)$, we have constraint

$$\sum_{c\in C(t)} \sum_{i\in M_j} \hat{R}_{jic}(\mathbf{x}) \ge L_j(t), \quad \forall j \in B, \tag{30}$$

ensuring that traffic demand is satisfied. Traffic demand is assumed to be known at the beginning of the epoch.

Remark 4. *Average raw throughput can be interpreted as a lower bound; better throughput could be obtained with additionally performing power control during an epoch. Also note that channel assignment granularity is in tens of seconds,*

[5]Whenever interference is observed, *observed average raw throughput* is calculated by $R_{jic}(\mathbf{x}) = \frac{w_c}{|M_j|}\log\left(1 + \frac{P_{jc}\bar{g}_{jic}d_{ji}^{-\alpha}}{\bar{I}_{ic}+N_0}\right)\mathsf{x}_{cj}$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCCN.2019.2925348, IEEE Transactions on Cognitive Communications and Networking

while resource block allocation is in the order of milliseconds (consider LTE for example).

We are interested in minimizing total spectrum usage while satisfying average traffic load requirements of BSs, i.e.

CHANNEL ASSIGNMENT PROBLEM

$$
\begin{aligned}
Z_{\text{IP}} := \min_{\mathbf{x}} & \sum_{c\in C(t)}\sum_{j\in B} w_c \mathsf{x}_{cj} \text{ subject to} \\
& \sum_{c\in C(t)}\sum_{i\in M_j} \hat{R}_{jic}(\mathbf{x}) \geq L_j(t), \quad \forall j \in B \\
& \sum_{j'\in B\setminus j} \bar{I}_{j'ic}\mathsf{x}_{cjj'} \leq I_{max}, \quad \forall c \in C(t), \forall j \in B, \forall i \in M_j \\
& \mathsf{x}_{cjj'} \leq \mathsf{x}_{cj}, \quad \forall c \in C(t), \forall j, j' \in B \\
& \mathsf{x}_{cj} + \mathsf{x}_{cj'} - \mathsf{x}_{cjj'} \leq 1, \quad \forall c \in C(t), \forall j, j' \in B
\end{aligned} \tag{31}
$$

We choose to minimize total spectrum usage as a means towards operators coexisting fairly in the shared spectrum setting considered. It is well-known that selfishness degrades spectrum efficiency in shared/unlicensed spectrum settings. As there is no way to identify the operators from the measured/predicted channel vacancy data, it is not possible to factor in fairness explicitly in the objective function of communication-free channel assignment optimization. So we implicitly seek to enforce fairness among operators in the above way.

With high traffic and inadequate resources, constraints in (30) may not be always satisfied. Therefore, we introduce traffic penalty in the objective of the channel assignment optimization problem which is given by

TRAFFIC PENALTY

$$
\begin{aligned}
Z(\eta) := \min_{\mathbf{x}} & \sum_{c\in C(t)}\sum_{j\in B} w_c \mathsf{x}_{cj} + \sum_{j\in B}\eta_j\Gamma_j(\mathbf{x}) \text{ subject to} \\
& \sum_{j'\in B\setminus j} \bar{I}_{j'ic} x_{cjj'} \leq I_{max}, \quad \forall c \in C(t), \forall j \in B, \forall i \in M_j \\
& \mathsf{x}_{cjj'} \leq \mathsf{x}_{cj}, \quad \forall c \in C(t), \forall j, j' \in B \\
& \mathsf{x}_{cj} + \mathsf{x}_{cj'} - \mathsf{x}_{cjj'} \leq 1, \quad \forall c \in C(t), \forall j, j' \in B
\end{aligned} \tag{32}
$$

where Γ_j and $\eta_j \geq 0$, $\forall j \in B$ are gradients and Lagrangian multipliers, respectively:

$$\Gamma_j(\mathbf{x}) = L_j(t) - \sum_{c\in C(t)}\sum_{i\in M_j}\hat{R}_{jic}(\mathbf{x}).$$

Subgradient Optimization: We utilize subgradient optimization method to find "good" values of Lagrangian multipliers. Let us denote by δ^l the *step-size* in iteration l given by $\delta^l = \xi^l(Z(\eta^l) - Z^*)/\sum_{j\in B}\Gamma_j^2(\tilde{\mathbf{x}}^l)$ where $\tilde{\mathbf{x}}^l$ denote the assignment variables that solve optimally Lagrangian relaxed problem (32) in iteration l; $\xi^{l+1} = \frac{1}{2}\xi^l$ if Z did not increase in last Q iterations, otherwise $\xi^{l+1} = \xi^l$ with parameters $n_{max} > 1$, and $0 < \xi^0 \leq 2$; and we have $Z^* = \min_{0\leq l^*\leq l} Z(\eta^{l^*}) - (1+l)^{-1}$. Pseudo-code is given in Algorithm 1.

Algorithm 1 OPTIMAL CHANNEL ASSIGNMENT

Input: $\{\bar{I}_{jic}\}_{\forall j,i,c}$, $\{L_j(t), M_j\}_{\forall j}$, n_{iter}
Initialization: Choose starting values $\eta^0 = (\eta_1^0, \eta_2^0, \ldots, \eta_{|B|}^0)$
Extrapolate $\{\hat{\bar{\nu}}_{ic}\}_{\forall i,c}$ using KEM, KEC or KEV
if Z_{IP} not exists **then**
  while $l \leq n_{\text{iter}}$ **do**
    Compute $Z(\eta^l)$ and $\{\tilde{\mathsf{x}}_{cj}^l\}_{\forall c,j}$ from (32)
    Compute subgradients $\Gamma_j(\tilde{\mathbf{x}}^l), \forall j \in B$
    if $\Gamma_j^l = 0, \forall j \in B$ **then**
      STOP, because the optimal value has been found
    end if
    Compute δ^l
    Compute $\eta_j^{l+1} = \max(0, \eta_j^l + \delta^l\Gamma_j^l)$, $\forall j \in B$
    $l = l + 1$
  end while
end if

B. Benchmark: Channel Assignments with an Inter-operator Coordination Protocol (CP)

For evaluating the performance of our above described communication-free scheme, we consider a protocol which involves explicit coordination among operators via communication between them. Specifically, (i) operators inform each other about their current channel assignments, (ii) utilizing this information and current state of channel interference, they perform channel assignments asynchronously as follows: each operator perform assignments and broadcast this to opponent operators. By this way, they make effort not to interfere heavily with each other. Such a protocol shall provide a fair comparison with our proposed communication-free proactive channel assignment scheme. Practicality aside, the other obvious downside of this baseline scheme is the overhead associated with coordination in terms of delay and communication overhead but we overlook this drawback by not accounting this overhead.

Algorithm 2 CHANNEL ASSIGNMENT WITH COORDINATION

for every operator sequentially **do**
  $C \leftarrow$ select channels not used by others
  while Traffic load is not satisfied **do**
    $C \leftarrow C \cup$ a channel used by others
    Perform channel assignments using Algorithm 1
    if all channels selected **then**
      Exit from while-block
    end if
  end while
  Inform other operators about used channels
end for

VI. NUMERICAL RESULTS

In this section, we do numerical simulations for better understanding interference prediction error performance of our different AR models (§IV) and comparison of the proposed Communication Free (CF) optimal channel assignment with the baseline coordination protocol (CP) approach. We consider that available channels do not change over time. In every Monte Carlo iteration, we generate random locations of mobiles, and assume that mobiles are associated with their

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCCN.2019.2925348, IEEE Transactions on Cognitive Communications and Networking

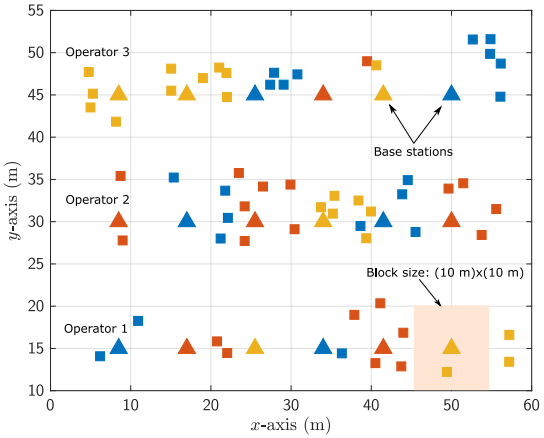


Fig. 5. Example scenario: Multi-operator deployment.

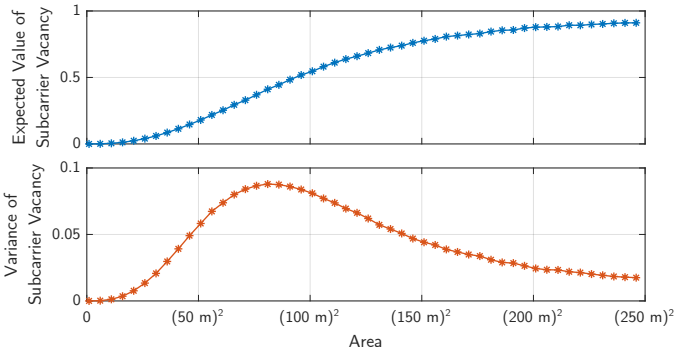


Fig. 6. Moments of subcarrier vacancy. Rayleigh fading: (Top) Expected value of subcarrier vacancy, (Bottom) Variance of subcarrier vacancy. There is only one interferer per area. Block size is kept unchanged.

nearest BS. Unless otherwise stated, we set system and channel parameters as follows: $P_s = 100/n_c$ mW, for every channel c and subcarrier s; Path loss exponent $\alpha = 3$, Rayleigh channel with variance 1; $\tau_e = -63/n_c$ dBm, $n_c = 1200$ subcarriers per every channel; Parameters in Algorithm 1: $n_{\text{iter}} = 100$, $n_{max} = 5$, $\xi^0 = 2$.

Locations of BSs and Mobiles: In every simulation, we divide the area into grids, and consider that a mobile is somewhere within a "block" (e.g., part of a building) and the corresponding BS at the center of that block (Figure 5). We randomly generate locations of mobiles located in the vicinity of its associated BS. We consider a (60 m)×(60 m) area with 3 operators, each has deployed 6 small BSs. Every operator serves 20 mobiles. So, totally there are 18 BSs and 60 mobiles in that area.

Epochs: We calculate mean of considered variables using Monte Carlo method on hourly basis ($n_e = 24$ epochs per day) by collecting data going back to 50 days, although we find that benefit of going back into the past beyond 20 days is marginal (see below).

Parameters of Kernel-based Extrapolation: For performing one step-ahead prediction, we use a sliding window of size $\Delta/2 + 1$ to estimate ρ. The first half part of the data $\Delta/2$ is used to compute ρ, and the last sample for performing one-step ahead prediction for different values of $\rho < \Delta/2 - 1$. We select the value of ρ which offers the lowest mean square error. The kernel used in extrapolating channel vacancy is Gaussian kernel $\kappa(v, v') = \exp(-|v - v'|^2/2\sigma_\kappa^2)$ where kernel variance σ_κ^2 is a critical parameter that determines the performance of the kernel, and we set it to $\sigma_\kappa^2 = 5$. We tried different kernel variances, and picked the aforementioned one because of its better performance. On the other hand, we found out that it is reasonable choosing $\Delta = 20$ (the values of $\Delta > 20$ did not significantly improve the performance).

Cost of Feedback: Duration of an epoch is $T = 60$ seconds. For every epoch, we set $n_d = 1000$ examples. We assume that there is no any loss or error when channel vacancy and location data is fed back to CC.

Traffic Load: For each mobile, we target 60 Mbps throughput; thus, if there are $|M_j|$ mobiles associated with BS j then, total traffic load becomes $L_j = 60|M_j|$ Mbps. Throughput per mobile is assumed to stay the same across epochs for simplicity. For every time instance, we consider that a mobile is active with probability 0.5.

A. Expected Value and Variance of Subcarrier Vacancy

In Section III-D, we have theoretical results about expected value and variance of subcarrier vacancy for a particular location grid of interferers. In this subsection, we aim at calculating related probabilities, expected value and variance by generating randomly locations. We assume that the mobile that is exposed to interference is at the center of the area, and the channel coefficient follows Rayleigh fading with unit variance.

In Figure 6, we plot the change of expected value and variance of subcarrier vacancy with increasing values of area. We assume that there exists one interferer on the area. Note that variance has a non-monotonic characteristic which has a pick around $(80\text{ m})^2$. It can be easily understandable since the interference strength will be high when area size is small and exactly opposite when area size is large. Therefore, interference strength will not fluctuate much in both cases which causes low variance. On the other hand, expected value of subcarrier vacancy must decrease when area size decreases since interference strength becomes high.

B. Error Performance of Kernel-based Extrapolation

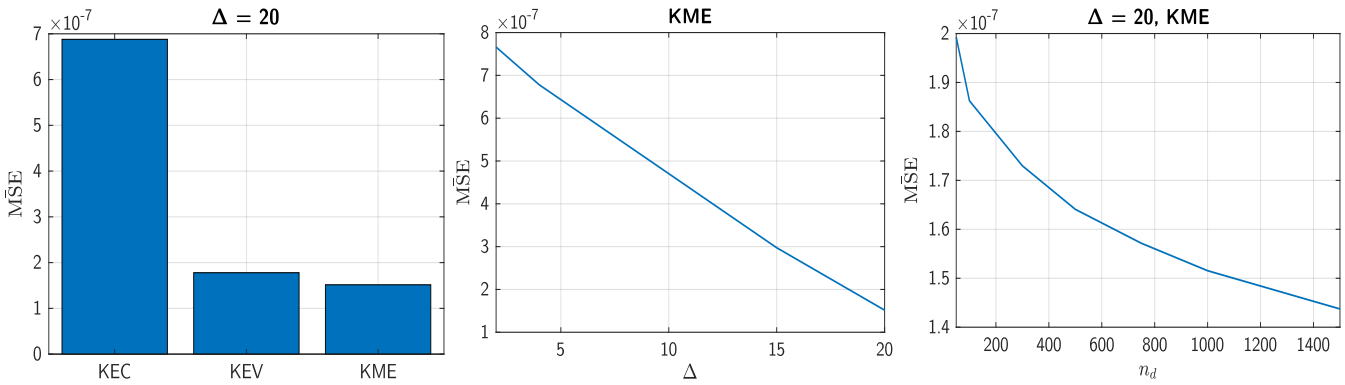
For any epoch t, mean square error is given by

$$\text{MSE}(t) = \frac{1}{50} \sum_{c \in C(t)} \sum_{i \in M} \sum_{D=1}^{50} |\bar{v}_{ci}(t - 24D) - \hat{\bar{v}}_{ci}(t - 24D)|^2$$

which captures the fact that t and D show the hour and day, respectively; further, we have $\bar{\text{MSE}} = \frac{1}{24} \sum_{t=1}^{24} \text{MSE}(t)$.

In Figure 7(a), we compare $\bar{\text{MSE}}$ performance of KEC, KEV and KME for $\Delta = 20$. KME has the best error performance among the three alternatives since it embeds all channel vacancy data. KEV is the second best performing alternative. It is however data efficient as it embeds only the expected value of channel vacancy. KEC performs worse among the three since the embedded error to RKHS may lead to degradation of performance. In particular, the underlying assumption in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCCN.2019.2925348, IEEE Transactions on Cognitive Communications and Networking



(a) MSE comparison of different learning techniques.

(b) MSE with respect to Δ.

(c) MSE with respect to n_d.

Fig. 7. MSE performance.

KEC that error is zero-mean normal random variable in its original space does not seem accurate.

Figure 7(b) plots MSE performance of KME with respect to increasing values of Δ. It shows the obvious advantage of using more data for extrapolation. On the other hand, Δ > 20 did not provide better performance in terms of spectrum efficiency, which is examined in the next subsection. We, however, note that the effect of Δ needs to be further explored in different scenarios and as such is an issue for future work. Figure 7(c) shows MSE performance of KME when n_d is increased. The result is not surprising to show that high volume of collected data is expected to demonstrate better error performance.

C. Spectrum Efficiency Performance

Here we study the spectrum usage and efficiency performance resulting from different extrapolation techniques (KME, KEV, KEC) that could be used with our proposed CF scheme. To show the value of channel vacancy forecasting, we also include the case of "no forecast", which uses the observed data from same period in previous day to inform its decisions. As a baseline, we consider an idealized CP scheme as outlined in §V.B.

In Figure 8, empirical CDF of total allocated spectrum for 50-day data is shown for different approaches. Clearly, CF with no forecasting yields the worst performance, indicating that simplistic extrapolation by relying on previous day data is ineffective. CF with no forecasting also results in the lowest spectrum efficiency compared to other approaches as shown in Figure 9, which depicts mean spectrum efficiency of three operators per epoch and its daily mean over 24 epochs.

There is also a direct relation between error performance of an extrapolation technique and the resulting spectrum efficiency. While KME has the best spectrum efficiency (even better than CP), KEC has the worst in line with its poor error performance (see Figure 7(a)). CF with KME and no-forecast, respectively, provide about 7.5 bps/Hz and 6 bps/Hz in terms of mean spectrum efficiency. This improvement in spectrum efficiency with KME translates to being able to support 900Mbps of traffic with 120 MHz of spectrum, compared to 720Mbps without forecasting. Moreover, in the considered scenario, there are 20 mobiles per operator; so, average traffic supported per mobile in case of KME and no-forecast would, respectively, be 45 Mbps and 36 Mbps, implying a 25% performance gain with a carefully chosen channel vacancy forecasting technique in the form of KME.

While CDF of spectrum allocation with CP is very similar to KME (except for Operator 2), its daily average performance is somewhat worse than CF with KME. This could be due to the asynchronous nature of operator channel assignment decisions which may degrade utilization of channels. Also note that the form of CP considered in the evaluations is idealized not accounting for the penalty due to latency and overhead associated with communication among operators, so the spectrum usage and efficiency performance results shown for CP are in fact on the most optimistic side.

D. Spectrum Allocation Characteristics of CF (KME)

We now examine the spectrum allocation characteristics of our proposed CF scheme using KME as the channel vacancy extrapolation technique (in light of results from the previous two subsections). We aim to demonstrate how our approach adapts the spectrum allocation in response to time varying traffic demands from different operators, and also show how it achieves fairness in spectrum allocation across operators. This experiment shows allocations at the granularity of channels (18 channels assumed to be available in total for all operators in this experiment).

In Figure 10, we depict, for each operator, epoch by epoch change of channel assignment ratio, which is basically the probability of assigning a channel to an operator in an epoch. We also show the total spectrum allocation for an operator as the weighted sum of available channels with the channel assignment ratios serving as weights.

Plots have been obtained by varying the effective traffic for each operator across epochs. Orange bars show the level of traffic. Note that spectrum is allocated only on demand, whenever traffic exists. These results demonstrates the behavior of our approach over time and shows that it balances traffic (from one or more operators) across available channels as indicated by the evolution of channel assignment ratio per channel

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCCN.2019.2925348, IEEE Transactions on Cognitive Communications and Networking

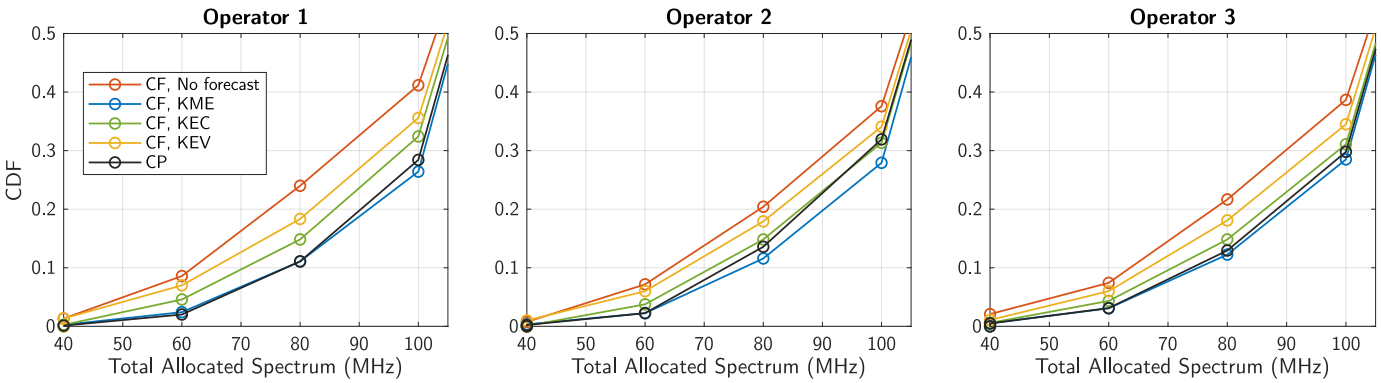


Fig. 8. Comparison of different schemes: CDF of total allocated spectrum. 50-day data has been used to generate empirical CDF.

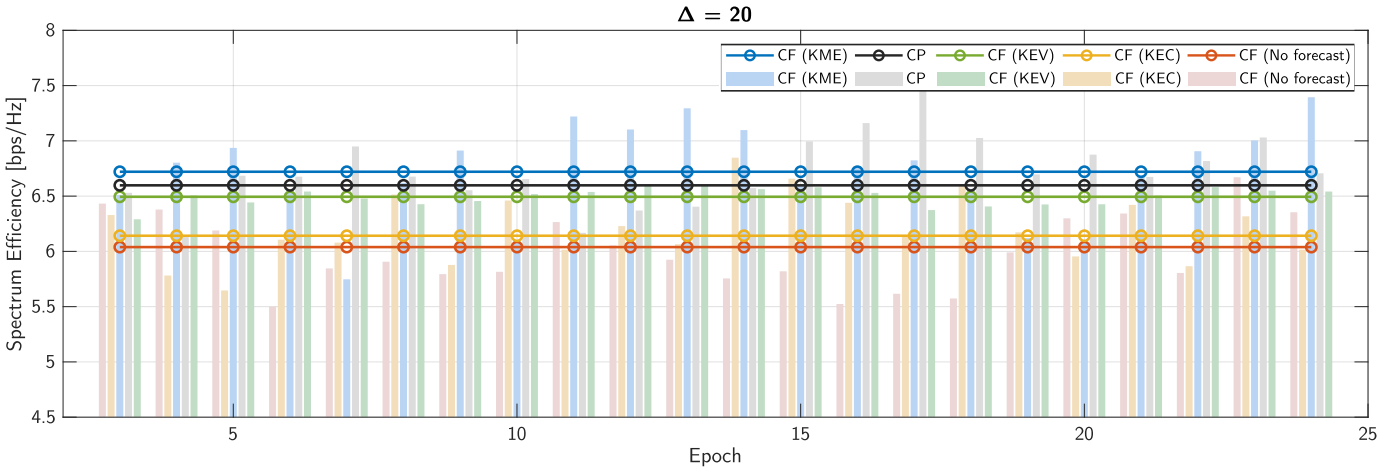


Fig. 9. (Bar) Mean spectrum efficiency per epoch. Mean has been calculated using a 50-day data of spectrum efficiency. Epochs are in hour-granularity. (Circle marked lines) Daily mean spectrum efficiency. CP = Coordination protocol, CF = Communication-free.

and per epoch. Moreover, these results show balanced/fair allocation of spectrum among competing operators with traffic demand.

E. Impact of Erroneous Feedback

Here we study the effect of erroneous feedback, which can be due to sensing errors (e.g., due to hardware limitations) or related to communication between the mobiles, BSs and controller of each operator. We consider a simple form of error called z-error, defined as the probability of receiving a vacancy data as zero instead of its actual value. Figure 11 plots the spectrum usage CDF with CF (KME) in presence of feedback error. Compared to the case where there is no feedback error, increasing z-error has the effect of increasing the total allocated spectrum. This can be explained by the fact that the lack of feedback (probabilistically in this experiment) inflates the spectrum occupancy from the perspective of each operator's controller, thereby increasing the overall amount of spectrum used but degrading the spectrum efficiency.

VII. Related Work

Various aspects of inter-operator spectrum sharing have received attention in the literature. [20] surveys the work on licensed spectrum sharing. In [21], the interaction of operators is modeled as a repeated non-cooperative game, and the utilities of the game are chosen to provide useful properties. Static and dynamic sharing is studied with cooperation and punishment states where once an operator deviates from the cooperation state, the punishment is everlasting, so that the operator suffers a net loss in revenue. [22] considers a two-stage game for investment and competition when spectrum is shared between a primary and secondary operators. The authors demonstrate that with their model only one secondary operator will invest and compete with the primary. In [23], two operator coexistence is studied where the first operator has only the macro-cell network architecture and does not possess small-cell network deployment, and the second operator has only the small-cell network architecture. Macro-cell network is benefited through offloading services offered by small-cell network, allowing macro-cell network to satisfy its users' capacity demands.

Coordination among secondary users in cognitive radio networks has been reasonably well studied in the literature with a number of MAC protocols proposed [24], [25], [26]. There are two distinct research directions within this literature, addressing two distinct problems. The first one focuses on opportunistic identification of the vacant of the spectrum by the secondary users; and then to transmit in them while guaranteeing that the primary users are minimally affected; refer to [27], [28], [29], [30] as representative literature. The

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCCN.2019.2925348, IEEE Transactions on Cognitive Communications and Networking

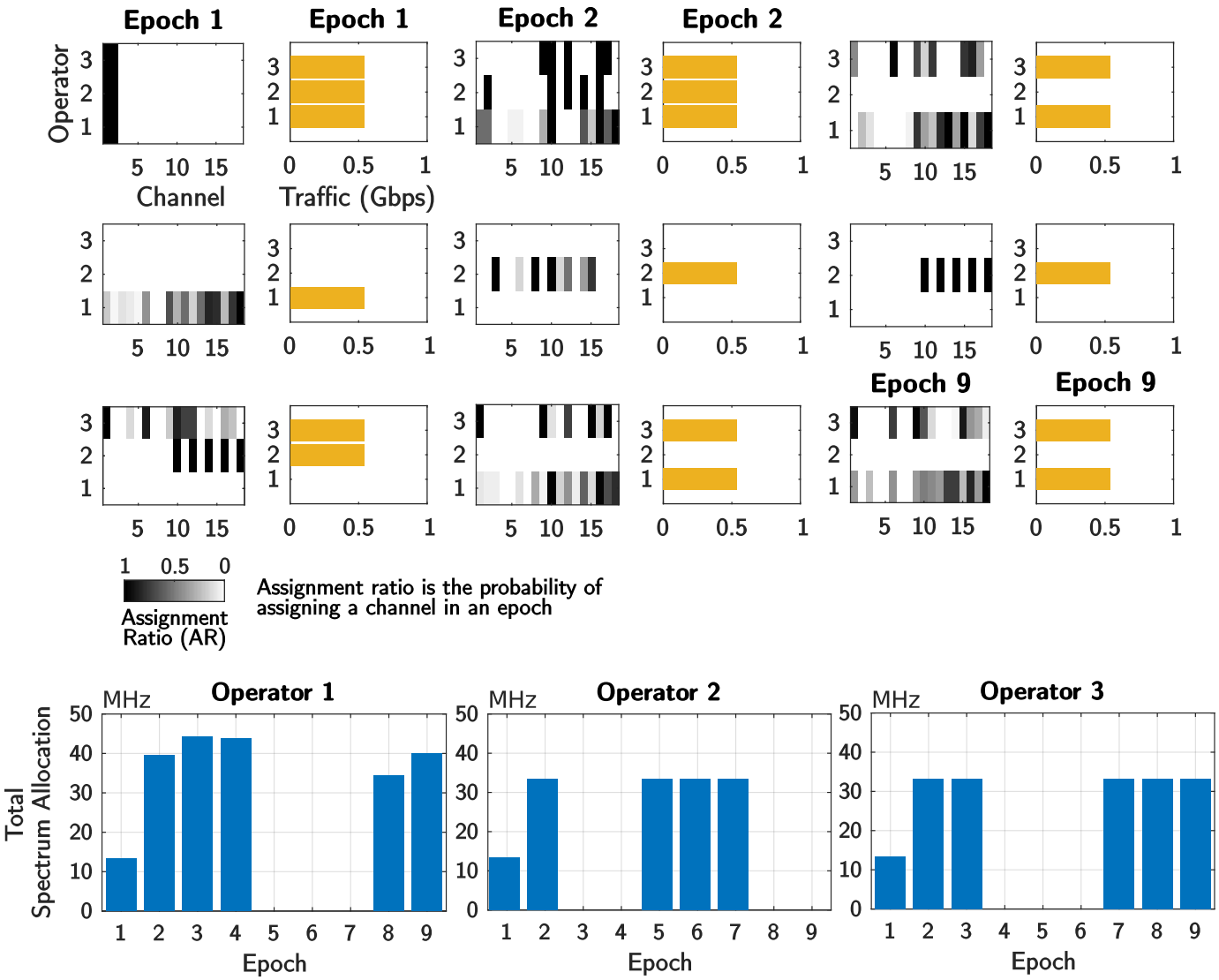


Fig. 10. Spectrum allocation behavior of the proposed CF (KME) scheme over time in a multi-operator scenario with operators having time-varying traffic demands.

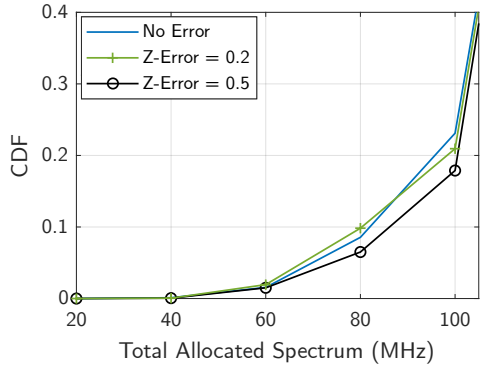


Fig. 11. Performance impact due to feedback error: CDF of total allocated spectrum. 50-day data has been used to generate empirical CDF.

second group of research is about developing protocols that enable secondary users to coordinate with each other when accessing the identified spectrum. The challenge is how users can/should optimally adapt their transmission strategies (including hopping among various frequency bands). Examples include [31], [32], [33], [6]. However, all of this work relies on a centralized coordinator to allocate the spectrum opportunities among the secondary users. More closer to our motivating problem context, [8] describes an end-to-end architecture for CBRS spectrum though it does not focus on spectrum sharing. But as part of their architecture, the authors mention coordination of secondary spectrum sharing among operators by the SAS, which as we stated at the outset limits the ability to perform dynamic and fine-grained spectrum sharing. Similarly in the work of [9], operators coordinate through the common Global Spectrum Controller (GSC).

An alternative approach to our problem is captured in the work of Chen and Huang [10] which targets the secondary spectrum sharing in the TV white space context. In particular, we consider the cooperative AP channel selection algorithm from [10], which involves each white space AP independently selecting a channel when tasked randomly by the database

and informs its decisions to other APs. This approach (with or without the database involvement) can be seen as an example of a distributed approach to our problem *with explicit coordination*. We compare it with ours and show that we can get similar or better performance without having to communicate with other operators. Another alternative approach in the coordination category is the work in [11], where the authors propose an inter-operator communication protocol to facilitate exchange of spectrum usage favors between operators. This is a coordination in a tit-for-tat spirit involving negotiations based on reciprocal altruism. Our approach is new and radically different from the aforementioned work in that it does not require any coordination (and negotiation) among operators but rather each operator learns about the others through measurements. Moreover, as the identities of "interferers" cannot be gleaned from the sensed interference information, inter-operator interaction cannot be modeled as a multi-agent game. We therefore take the interference forecasting approach and have the operators independently determine their channel assignments based on that information. In a very abstract sense, the work in [12] follows a similar paradigm to ours by sensing Wi-Fi spectrum usage through mobiles and then using that information for channel assignments. However the problem tackled there is much simpler and does not at all involve interference forecasting, the core contribution of this paper.

From a forecasting viewpoint, the authors in [37] use the classical time-series method of Holt-Winters for forecasting traffic of a slice in the 5G network slicing context to guide future admission control policies. In contrast, the interference forecasting problem we tackle is significantly harder as in our setting each operator needs to have an accurate notion of interference that is actually the consequence of continually changing actions (channel assignments) of other operators as well as other factors (time-varying channel conditions, user mobility, power control, etc.). There is literature on stochastic interference modeling aimed at studying spatio-temporal correlation and dynamics of interference, such as [38], [39], [40], [41], [42], [43], [44], [45], [46], [47]. A more recent work along these lines is in [48], which focuses on finite regions, as done in this paper, and shows that correlation of interference and outage are actually location-dependent. Although studying the correlation of interference over space and time could also be used for prediction of interference and outage over space and time, the above mentioned works do not focus on interference prediction methods. To the best of our knowledge, there is very limited work on interference prediction itself (e.g., [49], [50]) and certainly there does not exist any work that examines the problem at the depth and generality as our work. Leung [49] proposes a Kalman filter based approach for exploiting temporal correlation towards interference prediction for power control. In more recent work, the authors in [50] present a simple interference prediction technique that combines current interference measurement through pilot symbols and interferer traffic prediction, and uses it for slot selection. In [17], the order of AR model is one; in this work, we take into account the general case and solve corresponding least squares problem. [18] considers AR model in tensor product reproducing kernel Hilbert space. Authors show its performance against other AR models.

VIII. Conclusions

We have studied inter-operator spectrum sharing problem in shared spectrum small cell networks for which we devise a communication-free proactive optimal channel assignment scheme. As a key enabler of this scheme and core contribution of this work, we analyzed channel vacancy data which is defined over inter-operator interference, and by using kernel mean embedding technique, applied vector-valued regression for predicting time-varying probability distribution of channel vacancy and its expected value using various autoregressive (AR) models. We have compared our communication free scheme with coordination protocol and the case that does not involve any forecasting. Through extensive simulations, we showed that communication-free scheme performs at least as good or better than an idealized coordination protocol (with zero coordination overhead) and always outperforms the no forecasting alternative.

A key avenue for future work would be to investigate other potential applications of the proposed interference prediction approach for future mobile networks, as alluded to at the outset. Another issue would be to closely examine the extent to which other types of relatively simpler prediction problems in wireless communications (e.g., traffic, user mobility) can aid in the inteference prediction process.

Acknowledgment

This work was supported in part by The Leverhulme Trust. We would like to thank Uğur Özdemir and the anonymous reviewers for their valuable comments and suggestions.

References

[1] J. Zander and P. Mahonen, "Riding the data tsunami in the cloud: Myths and challenges in future wireless access," *IEEE Communications*, Mar 2013.

[2] FCC, "FCC Rule Making on 3.5 GHz Band/Citizens broadband radio service," Apr 2015.

[3] Ofcom, "3.8 GHz to 4.2 GHz Band: Opportunities for innovation," *White Paper*, Apr 2016.

[4] K. Mun, "CBRS: New shared spectrum enables flexible indoor and outdoor mobile solutions and new business models," *White Paper*, Mar 2017.

[5] M. Matinmikko et al., "Spectrum sharing using licensed shared access: The concept and its workflow for LTE-ADVANCED networks," *IEEE Wireless Communications*, Apr 2014.

[6] S. Rathinakumar and M. K. Marina, "GAVEL: Strategy-proof ascending bid auction for dynamic licensed shared access," *ACM MobiHoc*, 2016.

[7] MulteFire Alliance, "MulteFire release 1.0 technical paper," *White Paper*, Jan 2017.

[8] C. W. Kim, J. Ryoo and M. M. Buddhikot, "Design and implementation of an end-to-end architecture for 3.5 GHz shared spectrum," *IEEE DySPAN*, 2015.

[9] M. G. Kibria, G. P. Villardi, K. Ishizu, F. Kojima and H. Yano, "Resource allocation in shared spectrum access communications for operators with diverse service requirements," *EURASIP J. Adv. Signal Process.*, 2016.

[10] X. Chen and J. Huang, "Database-Assisted Distributed Spectrum Sharing," *IEEE Journal on Selected Areas in Communications*, Vol. 31, No. 11, Nov 2013.

[11] B. Singh et al., "Coordination protocol for inter-operator spectrum sharing in co-primary 5G small cell networks," *IEEE Communications*, Jul 2015.

[12] J. Shi et al., "A Walk on the Client Side: Monitoring Enterprise Wifi Networks Using Smartphone Channel Scans," *IEEE INFOCOM*, 2016.
[13] S. Yoon, L. E. Li, S. C. Liew, R. R. Choudhury, I. Rhee and K. Tan, "QuickSense: Fast and energy-efficient channel sensing for dynamic spectrum access networks," *IEEE INFOCOM*, 2013.
[14] A. Smola, A. Gretton, L. Song, B. Schölkopf, "A Hilbert space embedding for distributions," *in Hutter M., Servedio R.A., Takimoto E. (eds) Algorithmic Learning Theory. ALT 2007. Lecture Notes in Computer Science*, vol 4754. Springer, Berlin, Heidelberg
[15] L. Song, K. Fukumizu and A. Gretton, "Kernel embeddings of conditional distributions: A unified kernel framework for nonparametric enference in graphical models," *in IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 98-111, July 2013.
[16] M. Kallas, P. Honeine, C. Francis, H. Amoud, "Kernel autoregressive models using Yule-Walker equations," *Signal Processing*, vol. 93, no. 11, pp. 3053-3061, 2013.
[17] C. H. Lampert, "Predicting the future behavior of a time-varying probability distribution," *IEEE CVPR 2015*, 2015.
[18] E. A. Valencia, M. A. lvarez, "Short-term time series prediction using Hilbert space embeddings of autoregressive processes," *Neurocomputing*, vol. 266, pp. 595-605, 2017.
[19] H. Lian, "Nonlinear functional models for functional responses in reproducing kernel hilbert spaces," *The Canadian Journal of Statistics / La Revue Canadienne De Statistique*, vol. 35, no. 4, 2007, pp. 597-606.
[20] R. H. Tehrani, S. Vahid, D. Triantafyllopoulou, H. Lee and K. Moessner, "Licensed spectrum sharing schemes for mobile operators: A survey and outlook," in *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2591-2623, Fourthquarter 2016.
[21] F. Teng, D. Guo and M. L. Honig, "Sharing of unlicensed spectrum by strategic operators," in *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 3, pp. 668-679, March 2017.
[22] C. Liu, S. Fu and R. A. Berry, "Investing in shared spectrum," in *IEEE DySPAN 2017*, 2017.
[23] M. G. Kibria, G. P. Villardi, K. Nguyen, K. Ishizu and F. Kojima, "Heterogeneous networks in shared spectrum access communications," in *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 1, pp. 145-158, Jan. 2017.
[24] H. Wang, H. Qin, and L. Zhu, "A survey on MAC protocols for opportunistic spectrum access in cognitive radio networks," in *Int. Conf. Comput. Sci. Software Eng.*, 2008.
[25] T. V. Krishna and A. Dasa, "A survey on MAC protocols in OSA networks," *Comput. Networks*, vol. 53, no. 9, Jun. 2009.
[26] C. Cormio and K. R. Chowdhury, "A survey on MAC protocols for cognitive radio networks," *Ad Hoc Networks*, vol. 7, no. 7, Sep. 2009.
[27] J. Jia, Q. Zhang, and X. Shen, "HC-MAC: A hardware-constrained cognitive MAC for efficient spectrum management," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, Jan. 2008.
[28] Q. Zhao, L. Tong, A. Swami and Y. Chen, "Decentralized cognitive MAC for opportunisitic spectrum access in ad hoc networks: a POMDP framework," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, Apr. 2007.
[29] A. Hsu, D. Wei, C. Kuo, "A cognitive mac protocol using statistical channel allocation for wireless ad-hoc networks," in *Proc. Wireless Communications and Networking Conference (WCNC)*, 2007.
[30] S. Huang, X. Liu and Z. Ding, "Opportunistic spectrum access in cognitive radio networks," *IEEE INFOCOM* 2008.
[31] J. Huang, R. Berry, and M. Honig, "Auction-based spectrum sharing," *ACM Journal on Mobile Networks and Applications*, vol. 11, issue 3, Jun. 2006.
[32] F. Fu and M. van der Schaar, "Learning to compete for resources in wireless stochastic games," *IEEE Trans. Veh. Tech.*, vol. 58, no. 4, 2009.
[33] A. Fattahi, F. Fu, M. van der Schaar, F. Paganini, "Mechanism-based resource allocation for multimedia transmission over spectrum agile wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, Apr. 2007.
[34] J. Park and M. van der Schaar, "Cognitive MAC protocols using memory for distributed spectrum sharing under limited spectrum sensing," *IEEE Trans. Commun.*, vol. 59, no. 9, Sep. 2011.
[35] N. Taramas, G. C. Alexandropoulos and C. B. Papadias, "Opportunistic beamforming for secondary users in licensed shared access networks," *2014 6th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, Athens, 2014, pp. 526-529.
[36] N. I. Miridakis, T. A. Tsiftsis, G. C. Alexandropoulos and M. Debbah, "Simultaneous Spectrum Sensing and Data Reception for Cognitive Spatial Multiplexing Distributed Systems," in *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 3313-3327, May 2017.
[37] V. Sciancalepore *et al.*, "Mobile traffic forecasting for maximizing 5G network resource utilization," *IEEE INFOCOM*, 2017.
[38] R. K. Ganti and M. Haenggi, "Spatial and temporal correlation of the interference in aloha ad hoc networks," *IEEE Commun. Lett.*, vol. 13, no. 9, pp. 631-633, Sept. 2009.
[39] U. Schilcher, C. Bettstetter, and G. Brandner, "Temporal correlation of interference in wireless networks with rayleigh block fading," *IEEE Trans. Mobile Comput.*, vol. 11, no. 12, pp. 2109-2120, Dec. 2012. [
[40] M. Haenggi, "Diversity loss due to interference correlation," *IEEE Commun. Lett.*, vol. 16, no. 10, pp. 1600-1603, Oct. 2012.
[41] A. Crismani, S. Toumpis, U. Schilcher, G. Brandner, and C. Bettstetter, "Cooperative relaying under spatially and temporally correlated interference," in *IEEE Trans. Veh. Technol.*, 2015.
[42] U. Schilcher, S. Toumpis, A. Crismani, G. Brandner, and C. Bettstetter, "How does interference dynamics influence packet delivery in cooperative relaying?" in *Proc. ACM/IEEE Int. Conf. Modeling, Anal. and Simulation of Wireless and Mobile Syst.*, Nov. 2013, pp. 347-354.
[43] K. Gulati, R. Ganti, J. Andrews, B. Evans, and S. Srikanteswara, "Characterizing decentralized wireless networks with temporal correlation in the low outage regime," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3112-3125, Sept. 2012.
[44] Z. Gong and M. Haenggi, "The local delay in mobile poisson networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4766-4777, Sept. 2013.
[45] —-, "Temporal correlation of the interference in mobile random networks," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Kyoto, Japan, June 2011, pp. 1-5.
[46] —-, "Interference and outage in mobile random networks: Expectation, distribution, and correlation," *IEEE Trans. Mobile Comput.*, vol. 13, no. 2, pp. 337-349, 2014.
[47] Y. Zong, W. Zhang and M. Haenggi, "Managing interference correlation through random medium access," in *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, pp. 928-941, February 2014.
[48] K. Koufos and C. Dettmann, "Temporal correlation of interference and outage in mobile networks over one-dimensional finite regions," in *IEEE Transactions on Mobile Computing*, vol. 17, no. 2, Feb 2018.
[49] K. K. Leung, "Power control by interference prediction for broadband wireless packet networks," in *IEEE Transactions on Wireless Communications*, vol. 1, no. 2, pp. 256-265, April 2002.
[50] M. K. Atiq, U. Schilcher, J. F. Schmidt, and C. Bettstetter, "Semi-blind interference prediction in wireless networks", in *Proceedings of the 20th ACM International Conference on Modelling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM '17)*. ACM, New York, NY, USA, 2017.
[51] M. Ding et al., "On the Fundamental Characteristics of Ultra-Dense Small Cell Networks," in *IEEE Network*, Vol. 32, No. 3, May 2018.
[52] N. Liakopoulos et al., "Robust User Association for Ultra Dense Networks," *IEEE INFOCOM*, 2018.
[53] R. Li, C. Zhang, R. Stanica, F. Valois and P. Patras, "Learning Driven Mobility Control of Airborne Base Stations in Emergency Networks," in *Workshop on AI in Networks (WAIN)*, Dec 2018.
[54] F. Boccardi et al., "Spectrum Pooling in MmWave Networks: Opportunities, Challenges, and Enablers," in *IEEE Communications*, Vol. 54, No. 11, Nov 2016.
[55] H. Claussen, D. Lopez-Perez, L. Ho, R. Razavi and S. Kucera, "Small Cell Networks: Deployment, Management, and Optimization," Wiley-IEEE Press, Nov 2017.
[56] P. Billingsley, *Probability and Measure* (2nd ed.). Wiley, 1986, pp. 369.

APPENDIX

A. Proof of Theorem 1

Lindeberg's variant of central limit theorem only requires that random variables must have finite expected value and variance but may not be identically distributed [56]. Thus, if sequence of independent random variables $X_{ic,s}$, for every $s = 1, \ldots, n_c$ satisfies Lindeberg's condition:

$$\lim_{n_c \to \infty} \frac{1}{S_{n_c}^2} \sum_{s=1}^{n_c} \mathbb{E}\left[(X'_{ic,s})^2 \cdot \mathbf{1}\left\{|X'_{ic,s}| \geq \varepsilon S_{n_c}\right\}\right] = 0 \qquad (33)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCCN.2019.2925348, IEEE Transactions on Cognitive Communications and Networking

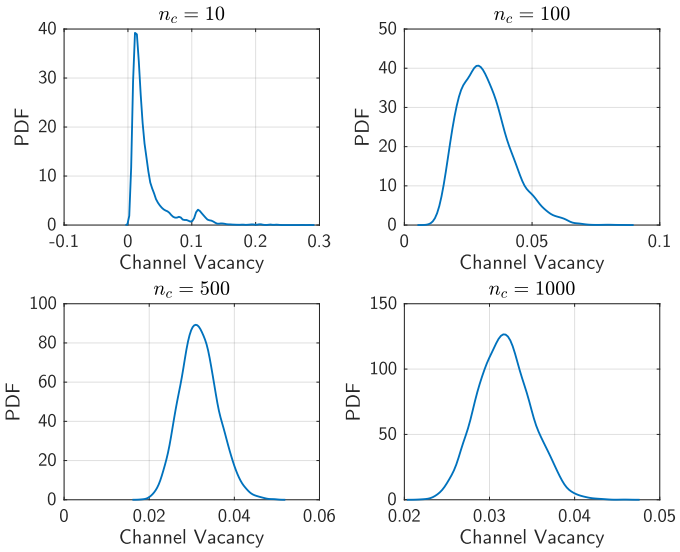


Fig. 12. PDF of channel vacancy for increasing values of subcarriers.

for all $\varepsilon > 0$, where $X'_{ic,s} = X_{ic,s} - \bar{X}_{ic,s}$ and $S^2_{n_c} = \sum_{s=1}^{n_c} \sigma^2_{X_{ic,s}}$, then central limit theorem holds. Let us rewrite the following:

$$\mathbb{E}\left[(X'_{ic,s})^2 \cdot \mathbf{1}\left\{|X'_{ic,s}| \geq \varepsilon S_{n_c}\right\}\right] = \int_{-\infty}^{-\varepsilon S_{n_c}} x^2 \mathsf{pdf}_{X'_{ic,s}}(x)dx + \int_{\varepsilon S_{n_c}}^{\infty} x^2 \mathsf{pdf}_{X'_{ic,s}}(x)dx = \underbrace{\int_{0}^{\bar{X}_{ic,s}-\varepsilon S_{n_c}} (x-\bar{X}_{ic,s})^2 \mathsf{pdf}_{X_{ic,s}}(x)dx}_{V_s} + \underbrace{\int_{\bar{X}_{ic,s}+\varepsilon S_{n_c}}^{1} (x-\bar{X}_{ic,s})^2 \mathsf{pdf}_{X_{ic,s}}(x)dx}_{W_s}.$$

Then left-side of (33) can be given by

$$\lim_{n_c\to\infty} \frac{1}{S^2_{n_c}} \sum_{s=1}^{n_c} (V_s + W_s) = \lim_{n_c\to\infty} \frac{1}{S^2_{n_c}} \sum_{s=1}^{n_c} V_s + \lim_{n_c\to\infty} \frac{1}{S^2_{n_c}} \sum_{s=1}^{n_c} W_s.$$

Note that for every $\varepsilon > 0$, we have $\sum_{s=1}^{n_c} V_s < S^2_{n_c}$ and $\sum_{s=1}^{n_c} W_s < S^2_{n_c}$ if $n_c \to \infty$. Thus, the condition in (33) holds with

$$\lim_{n_c\to\infty} \frac{1}{S^2_{n_c}} \sum_{s=1}^{n_c} V_s = 0 \text{ and } \lim_{n_c\to\infty} \frac{1}{S^2_{n_c}} \sum_{s=1}^{n_c} W_s = 0.$$

In simulations, we have results verifying the fact proved that probability density of channel vacancy converges to normal distribution. In Figure 12, we depict PDF of channel vacancy for increasing values of subcarriers. Channel fading coefficient is assumed to follow Rayleigh distribution. We generated randomly locations of 10 small BSs on an area of (100 m) × (100 m) for an example scenario. We assumed τ_e = 63 dBm, power per subcarrier 80 μW.

B. Proof of Theorem 2

Note that the solution of (16) can be found by solving the following least-squares functional for every $r = 1, \ldots, \rho$:

$$\min_{\mathbf{\Lambda}_r} \sum_{z=\rho}^{\Delta-1} \left\|\bar{\mu}^r_{z+1} - \mathbf{\Lambda}_r \hat{\mu}_{z-r+1}\right\|^2_{\mathscr{H}} + \lambda \|\mathbf{\Lambda}_r\|^2_{\mathscr{L}}, \quad (r = 1, \ldots, \rho) \tag{34}$$

where $\bar{\mu}^r_{z+1} = \hat{\mu}_{z+1} - \sum_{r'=1,r'\neq r}^{\rho} \Lambda_{r'} \hat{\mu}_{z-r'+1}$. The operator minimizing (34) is well-known [17], [19]:

$$\Lambda_r = \sum_{z=\rho}^{\Delta-1} \left(\hat{\mu}_{z+1} - \sum_{r'=1,r'\neq r}^{\rho} \Lambda_{r'} \hat{\mu}_{z-r'+1}\right) \sum_{z'=\rho}^{\Delta-1} Q_{z,z'} \hat{\mu}^\top_{z'-r+1}, \quad (r = 1, \ldots, \rho) \tag{35}$$

where $Q_{z,z'} \in \mathbf{Q} = (\mathbf{K} + \lambda \mathbf{I})^{-1}$, and entries of *kernel matrix* $\mathbf{K} \in \mathbb{R}^{(\Delta-1)\times(\Delta-1)}$ are given by $K_{z,z'} = \langle \mu_z, \mu_{z'} \rangle_{\mathscr{H}}$:

$$\langle \mu_z, \mu_{z'} \rangle_{\mathscr{H}} = \frac{1}{n_d^2} \sum_{k=1}^{n_d} \sum_{k'=1}^{n_d} \kappa\left(v^{k,z}, v^{k',z'}\right). \tag{36}$$

Solution in (35) creates a matrix system of equations denoted by $\mathbf{M}\mathbf{\Lambda}^\top = \mathbf{m}^\top$ which can give us needed linear operators, i.e.

$$\hat{\mathbf{\Lambda}} = [\mathbf{M}^{-1}]^\top \mathbf{m} \tag{37}$$

where $\hat{\mathbf{\Lambda}} = [\hat{\mathbf{\Lambda}}_1, \ldots, \hat{\mathbf{\Lambda}}_\rho]^\top$. Actually, we do not need to know explicitly $\hat{\mathbf{\Lambda}}$ to calculate expected value of any function f. On the other hand, $\mathbf{M} \in \mathbb{R}^{\rho\times\rho}$ and $\mathbf{m} \in \mathbb{R}^{\rho n_h \times n_h}$ where n_h is number of dimensions when mapping to RKHS which we do not know explicitly. Further, we define the following:

$$\mathbf{M} = \begin{bmatrix} 1 & M_{1,2} & \ldots & M_{1,\rho} \\ M_{2,1} & 1 & \ldots & M_{2,\rho} \\ \vdots & \vdots & \ddots & \vdots \\ M_{\rho,1} & M_{\rho,2} & \ldots & 1 \end{bmatrix}, \mathbf{m} = \begin{bmatrix} \tilde{\mathbf{m}}_1 \\ \tilde{\mathbf{m}}_2 \\ \vdots \\ \tilde{\mathbf{m}}_\rho \end{bmatrix}$$

where for all $r, r' = 1, \ldots, \rho$

$$M_{r,r'} = \left[\sum_{\substack{z=\rho \\ z'=\rho}}^{\Delta-1} Q_{z,z'} \hat{\mu}_{z-r'+1} \hat{\mu}^\top_{z'-r+1}\right]^\top = \sum_{\substack{z=\rho \\ z'=\rho}}^{\Delta-1} Q_{z,z'} \langle \hat{\mu}_{z-r'+1}, \hat{\mu}_{z'-r+1} \rangle$$

and $\tilde{\mathbf{m}}_r = \sum_{\substack{z=\rho \\ z'=\rho}}^{\Delta-1} Q_{z,z'} \hat{\mu}_{z+1} \hat{\mu}^\top_{z'-r+1}$.

Thus, any linear operator $\hat{\mathbf{\Lambda}}_r$ can be given by

$$\hat{\mathbf{\Lambda}}_r = \sum_{r'=1}^{\rho} \Upsilon_{r,r'} \tilde{\mathbf{m}}_{r'} = \sum_{\substack{z=\rho \\ z'=\rho}}^{\Delta-1} Q_{z,z'} \hat{\mu}_{z+1} \sum_{r'=1}^{\rho} \Upsilon_{r,r'} \hat{\mu}^\top_{z'-r'+1}.$$

where $\Upsilon = [\mathbf{M}^{-1}]^\top$.